
Brian Garrett
¿QUÉ ES ESO
LLAMADO METAFÍSICA?
Alianza Editorial

Brian Garrett

# ¿Qué es eso llamado metafísica?

Traducción española de Fabio Morales

Alianza Editorial

Edición electrónica, 2014
www.alianzaeditorial.es


Juan Ignacio Luca de Tena, 15. 28027 Madrid
ISBN: 978-84-206-8912-8
Edición en versión digital 2014

*Para Elizabeth Remonde Garrett,*
*con amor.*

«El objeto de una educación liberal es producir la perfección de la mente; [...] y mientras que una educación profesional es un adiestramiento para el trabajo, una educación liberal es un adiestramiento (si es que podemos denominarlo así) para el ocio; nos enseña, para emplear la expresión de Pattison, el arte de vivir: instruye al hombre en el modo de habitar y desenvolverse en el mundo y contemplarlo como corresponde a un ser civilizado.»

John Sparrow, *Mark Pattison and the Idea of a University*
(Cambridge: Cambridge University Press), 1966

«Omítanse las palabras innecesarias.»

W. Strunk y E. B. White, *The Elements of Style*
(Toronto: Macmillan), 1970

# Índice

# Prefacio

Comencé a escribir este libro durante un permiso sabático que me concedió la Universidad Nacional de Australia (ANU) en la primera mitad de 2005. Agradezco a la ANU la dispensa de enseñar durante ese periodo. En posteriores seminarios que sostuve en la ANU, mientras trabajaba en borradores de capítulos, recibí valiosos comentarios de mis colegas. Gracias, en particular, a Jeremy Shearmur y Peter Roeper, así como a Havi Carel y Udo Thiel. Gracias, también, a los estudiantes de postgrado que asistieron a los seminarios: David Wall, Peter Grundy, Peter Eldridge-Smith, Luc Small, Paul Miller y Matt Cox.

Quiero agradecer a Tony Bruce, director de la Colección de Filosofía de la editorial Routledge, sus ánimos; a Priyanka Pathak, la directora de Proyectos, sus recomendaciones y comentarios, y a Tim Crane, que me pusiera en contacto con Tony. Introduje una serie de enmiendas como resultado de las observaciones críticas de tres lectores de Routledge. Además, mi antiguo alumno Robert Nichols desplegó su sagaz pluma editorial sobre todo el texto; y Thomas Mautner me permitió beneficiarme de su amplio conocimiento de la historia de la filosofía.

En un intento por hacer el libro tan fácil de manejar como sea posible, Routledge ha sugerido la adopción de un formato en el que cada capítulo incluya recuadros para conceptos, perfiles de filósofos, citas clave, glosas de términos, cuestionarios para el estudio, referencias a lecturas adicionales y recursos de Internet. Espero que el lector los encuentre útiles. Pero si, a pesar de todo, algunas palabras o frases clave siguieran estando confusas, recomiendo encarecidamente la ayuda inestimable que puede prestar el *Dictionary of Philosophy* de Thomas Mautner (2.ª edición, Londres: Penguin, 2005).

# Introducción

Cuando Andrónico de Rodas editó las obras de Aristóteles en el siglo I a.C., colocó los libros que tenían que ver con la «filosofía primera» después de los de física, dando así lugar a que la posteridad los conociera más tarde bajo el nombre de *meta*física. Aristóteles describió su «filosofía primera» como el estudio del ser *en cuanto* ser. Esta caracterización, aunque críptica, logra captar un aspecto medular de la metafísica: su búsqueda de generalidad. El metafísico se esfuerza en dilucidar los rasgos más generales y ubicuos de la realidad; por ejemplo: la existencia, el espacio, el tiempo, la causalidad o las nociones de objeto y de propiedad. Pero la indagación metafísica no siempre es absolutamente general. También incluye el estudio de temas más específicos, como el libre albedrío y la identidad personal.

Al realizar sus investigaciones, los estudiosos de la metafísica se valen de una plétora de datos y técnicas: el pensamiento riguroso, el análisis conceptual, los estudios empíricos y los métodos formales de la lógica y de la matemática. Los resultados empíricos y las técnicas formales resultan más útiles para unos temas metafísicos que para otros. Por ejemplo, las conclusiones empíricas son importantes para la discusión sobre la naturaleza del espacio y del tiempo; y las técnicas formales lo son sobre todo para el debate entre el realismo y el antirrealismo. Y, sin embargo, las controversias en metafísica rara vez son zanjadas recurriendo únicamente a resultados empíricos o a técnicas formales de la lógica y de las matemáticas. Los asuntos metafísicos han de verse como si estuvieran en un espectro en el que unos se hallan situados en el extremo más abstracto y a priori, mientras que otros tienden al extremo donde la ciencia empírica cobra más protagonismo.

La mejor forma de captar la naturaleza de la metafísica consiste en sumergirse en las disputas metafísicas particulares. En esta obra he tratado de presentar los temas centrales de la metafísica de una manera clara y comprensible, despojada de los detalles que tienden a congestionar las revistas filosóficas especializadas. Frecuentemente tomo postura en puntos concretos. Pero, cuando lo hago, siempre aduzco mis razones, y el lector queda en libertad para disentir. En el resto de esta Introducción esbozaré los aspectos principales tratados en cada capítulo.

En el capítulo 1 examinamos los tres argumentos tradicionales a favor de la existencia de Dios (el ontológico, el cosmológico y el teleológico) y un argumento, bastante conocido, en contra de su existencia (el argumento del mal). La discusión introduce y recurre a nociones modales (las de posibilidad y necesidad) y, en especial, a la noción de un ser necesario.

En el capítulo 2 pasamos revista a varios puntos relacionados con la existencia. J. L. Austin dijo alguna vez, medio en broma, que la existencia era «como la respiración, sólo que más silenciosa». Es difícil entender exactamente lo que quiso decir, pero aquí, por lo pronto, echaremos un vistazo al tema del realismo modal (la idea de que otros mundos y entidades posibles existen exactamente en el mismo sentido en que lo hacen nuestro mundo y sus habitantes); examinaremos algunas posibles respuestas a la vieja pregunta «¿por qué hay algo en lugar de la nada?»; consideraremos la cuestión de si hay objetos no existentes, y, finalmente, abordaremos, por qué no, la interrogante de si la existencia, al igual que la respiración, es una propiedad de los objetos ordinarios. Nuestra discusión recurrirá un poco a la lógica elemental y pondrá de manifiesto varias conexiones entre estos cuatro tópicos.

El capítulo 3 se ocupa de un asunto metafísico típico y muy antiguo: el problema de los universales. ¿Es necesario concebir las propiedades como universales (idénticas en todas sus instancias) para explicar la naturaleza de las propiedades y las (supuestas) diferencias entre las propiedades y los particulares? Los realistas responden que sí; los nominalistas, que no. También examinamos una versión del nominalismo bastante reciente —la teoría de los tropos—, la cual subvierte el paradigma tradicional y concibe las propiedades como si fueran particulares.

En el capítulo 4 nos enfocamos en la causalidad. Después de analizar varias cuestiones preliminares —como, por ejemplo, los tipos de entidades que pueden ser causas o efectos, la direccionalidad temporal de la causalidad o las propiedades lógicas de la relación causal—, planteamos la pregunta metafísica fundamental sobre la naturaleza de la relación de causalidad: ¿Qué relación entre dos fenómenos hace que uno de ellos sea causa del otro? Analizamos dos respuestas reduccionistas a esta pregunta: la de Hume, que creía que la causalidad consiste en una conjunción constante, y la de Lewis, que la concebía como una dependencia contrafáctica. También examinamos una respuesta no reduccionista: la idea de que la causalidad es una especie irreductible de necesidad natural.

El capítulo 5 versa sobre el tiempo. Como J. M. E. McTaggart señaló a comienzos del siglo XX, las localizaciones en el tiempo pueden ser concebidas de dos maneras bastante divergentes entre sí. Los eventos pueden ser colocados en la serie B (y ordenados por medio de las relaciones anterior-a y posterior-a), y los eventos puede ser colocados en la serie A (y ordenados como pasados, presentes o futuros). Las posiciones de la serie B son permanentes; mientras que las de la serie A están cambiando continuamente. El teórico A considera las determinaciones de la serie A fundamentales para el concepto de tiempo, mientras que el teórico B considera las determinaciones de la serie B las fundamentales. McTaggart creyó que tenía objeciones decisivas a ambas teorías, y no vaciló en concluir que el tiempo era irreal. Pocos han seguido su heroica senda, pero el debate entre los teóricos de los tipos A y B, a menudo científicamente documentado, sigue vivo.

En el capítulo 6 abordamos algunos puntos secundarios sobre el tiempo. ¿Es un argumento a favor de la teoría A sobre el tiempo el que tengamos ciertas actitudes temporalmente prefiguradas (como cuando, por ejemplo, nos sentimos aliviados porque una experiencia dolorosa ha pasado, y no lo hacemos, en cambio, cuando está por suceder)? ¿Es, como muchos filósofos han sostenido, una verdad necesaria que no habría tiempo si no hubiera cambio? Finalmente, ¿es posible viajar hacia el pasado?

El capítulo 7 examina algunos desafíos a la creencia de sentido común de que poseemos libre albedrío. El fatalista, basado en razones puramente lógicas, arguye que la libertad es una ilusión: no podemos hacer sino lo que efectivamente hacemos. Los así llamados deterministas radicales alegan que el libre albedrío y el determinismo son incompatibles y que, como el determinismo es verdadero, careceríamos de libre albedrío. (Sometemos a escrutinio los argumentos del fatalista y los del determinista radical.) Otros filósofos, finalmente, sostienen que el libre albedrío exige un tipo de autodeterminación que es lógicamente imposible de satisfacer, y que, por lo tanto, la noción de libre albedrío es incoherente.

En el capítulo 8 trataremos de dilucidar qué es lo que caracteriza a las personas como tales, y en qué consiste ser la misma persona a través del tiempo. La metodología empleada descansa aquí en gran medida en experimentos mentales imaginarios y en juicios intuitivos sobre quién sigue siendo quién a lo largo de tales experimentos mentales. Este método posee sus limitaciones, pero permite decidir entre teorías rivales de la identidad personal. Al igual que sucedía con la causalidad, las teorías de identidad personal pueden ser divididas en variantes reduccionistas y no reduccionistas. Se someten a crítica pormenorizada todas las teorías reduccionistas convencionales, y se propone como alternativa una teoría no reduccionista: la concepción simple.

El capítulo final se ocupa de una versión moderna de un viejo debate. La disputa del realismo contra el idealismo es una de las más antiguas de la filosofía. En la década de 1960, el filósofo de Oxford Michael Dummett propuso una nueva forma de concebirla. Apoyándose en diversas ideas del segundo

Wittgenstein relativas a la naturaleza pública del significado, Dummett sugirió que se entendiera el realismo como una manera de concebir el significado. Un realista sobre los F (otras mentes, el pasado, la matemática, etc.) es alguien que sostiene que, en virtud del significado de los enunciados-F, los enunciados-F son definidamente verdaderos o falsos, independientemente de si alguna vez podemos llegar o no a conocer su valor de verdad. Un antirrealista sobre los F niega tal determinabilidad veritativa. En el capítulo 9 consideramos la motivación que subyace al antirrealismo, la naturaleza de sus tesis, y analizamos un argumento técnico que incluso pone en duda su coherencia.

# 1. Dios

## Introducción

Una de las preguntas metafísicas más antiguas es: ¿Existe Dios? Aquí la abordaremos entendiendo a «Dios» en el sentido filosófico de la tradición, es decir, como un ser **omnipotente**, **omnisciente** y completamente bueno. En este capítulo analizamos tres de los argumentos mejor conocidos a favor de la existencia de Dios, así como un argumento famoso en contra de su existencia. Los tres argumentos a favor de la existencia de Dios suelen ser conocidos como el argumento ontológico, el argumento cosmológico y el argumento teleológico, aunque existen muchas variantes de cada uno. El argumento contra la existencia de Dios que examinaré es el argumento sobre el mal.

El argumento ontológico intenta probar la existencia de Dios recurriendo únicamente a la razón. Se trata de un argumento completamente **a priori**. El meollo del mismo es que basta con aprehender el concepto o la idea de Dios, junto con aquello que este concepto o idea implica, para demostrar que existe Dios. Por consiguiente, el argumento ontológico se propone, partiendo de premisas cognoscibles a priori, ser una prueba deductivamente válida de la existencia de Dios. Pretende ser tan sólido y convincente como cualquier prueba usada típicamente en la lógica o en las matemáticas.

El argumento cosmológico y el teleológico no están basados en premisas cognoscibles a priori, sino en premisas **contingentes** o empíricas. El argumento cosmológico parte del hecho de que el universo que nos rodea existe. Y prosigue así: puesto que la existencia del universo es contingente (en lugar de

algo, habría podido no haber nada), debe existir un **ser necesario,** o no contingente, que lo creó.

El argumento teleológico (o argumento del designio) también procede de una premisa empírica: no la premisa de que existe el universo, sino la más específica de que el universo contiene entidades maravillosamente complejas, tales como flores, ojos y cerebros, cuya existencia, se alega, respalda la hipótesis de un creador benevolente y amoroso. Sería sumamente improbable que entidades estupendas como ésas apareciesen por casualidad; son fruto, pues, del designio. Obsérvese cómo este argumento no tiene carácter deductivo. No se está aseverando que la existencia de estructuras complejas como los ojos y los cerebros implique lógicamente la existencia de Dios, sino que la existencia de tales estructuras hace razonable creer que existe Dios. La postulación de un creador divino es lo que mejor explicaría, de acuerdo con esto, la existencia de tales estructuras.

Un argumento importante contra la existencia de Dios es el argumento del mal. Niega la existencia de un Dios concebido como un ser amoroso, omnisciente y todopoderoso. El argumento del mal asevera que la existencia del mal es incompatible con la existencia de Dios. La versión menos radical del argumento del mal, basada en la evidencia, se limita a afirmar que la existencia de Dios proporciona evidencia contra la existencia de Dios, y concluye que, dada la proliferación del mal, sería irracional creer en un Dios amoroso.

## El argumento ontológico

A lo largo de la historia de la filosofía han sido muy numerosas las versiones diferentes del argumento ontológico; pero la primera, y más discutida, es la que presentó en el siglo XI d.C. san Anselmo, arzobispo de Canterbury, en su *Proslogio*. He aquí un párrafo crucial, que nos permite seguir su razonamiento:

> Por lo tanto, incluso un necio está convencido de que aquello que es tal que nada mayor que ello puede ser concebido se encuentra en el entendimiento; puesto que cuando oye esto, lo entiende; y todo aquello que se puede entender está en el entendimiento. Y ciertamente, aquello que es tal que nada mayor que ello puede ser concebido no puede hallarse únicamente en el entendimiento. Pues si estuviera [...] sólo en el entendimiento, podría concebirse también en la realidad, que es algo mayor. Por lo tanto, si lo que es tal que nada mayor que ello puede ser concebido estuviera sólo en el entendimiento, entonces lo que es tal que nada mayor que ello puede ser concebido sería lo mismo que aquello que es tal que algo mayor que ello sí puede ser concebido[1].

Una reconstrucción posible de lo dicho en este pasaje es la siguiente:

(1) Dios es tal que nada mayor que Él puede ser concebido.
(2) Dios existe o bien sólo en el entendimiento, o existe tanto en el entendimiento como en la realidad.

(3) Si Dios existiese sólo en el entendimiento, podría concebirse un ser mayor, a saber, un ser que tuviera todas las cualidades de Dios y que además existiera tanto en el entendimiento como en la realidad.
(4) Pero Dios es tal que nada mayor que Él puede ser concebido ([1]).
(5) Por lo tanto, Dios no puede existir sólo en el entendimiento (por [3] y [4]).
(6) Por lo tanto, Dios existe tanto en el entendimiento como en la realidad (por [2] y [5]).
(7) Por lo tanto, Dios existe (en la realidad) (por [6]).

### San Anselmo (1033-1109)

San Anselmo nació en Aosta, Italia. Se convirtió en monje y fue nombrado más tarde arzobispo de Canterbury. Teólogo y filósofo a la vez, es famoso por haber ideado la primera versión del argumento teológico a favor de la existencia de Dios. La fe religiosa de san Anselmo no descansaba sobre esta prueba; simplemente quería poner de manifiesto la existencia y la naturaleza de Dios. Como dijo en cierta ocasión: «No trato de entender para poder tener fe, sino que tengo fe para poder entender». Su argumento ontológico tuvo una acogida desigual: santo Tomás de Aquino y Kant lo rechazaron, mientras que Duns Escoto y Descartes ofrecieron incluso sendas versiones de él. Aunque el argumento tiene hoy pocos seguidores, no existe consenso acerca de qué es lo que está errado en él.

La premisa (1) pretende ser una verdad puramente definicional. Para san Anselmo, la palabra «Dios» significa (entre otras cosas) «algo tal que nada mayor que ello puede ser concebido», así como «triángulo» significa «figura de tres lados y tres ángulos» y «soltera» significa «mujer no casada». Así pues, el necio no podría coherentemente negar que Dios sea tal que nada mayor que Él puede ser concebido, como tampoco podría negar coherentemente que los triángulos tengan tres lados o que las solteras no estén casadas.

La premisa (2) se entiende como un truismo, y como una instancia de la verdad, supuestamente general, siguiente: para cualquier F que haya sido concebido, o bien F existe únicamente en el entendimiento, o bien existe tanto en el entendimiento como en la realidad. Así, por ejemplo, los unicornios y los dragones existirían sólo en el entendimiento, mientras que los hombres y los caballos existirían tanto en el entendimiento como en la realidad.

La premisa (3) sigue la siguiente línea de pensamiento. Supongamos que consideramos a dos seres que son del todo semejantes en sus propiedades, salvo en el hecho de que el primero existe tan sólo en el entendimiento mientras que el segundo existe tanto en el entendimiento como en la reali-

dad. Entonces, el segundo ser será mayor que el primero: el existir en la realidad es una propiedad que hace que algo sea mayor. Este principio, tomado junto con la premisa (1), da lugar a la premisa (3). Dados (1), (2) y (3), se sigue fácilmente (7).

¿Cómo podríamos criticar este argumento? Para empezar, no deberíamos presuponer que todas las definiciones o estipulaciones son coherentes. Algunas no lo son. Por ejemplo, yo podría tratar de definir «meganúmero» de la manera siguiente:

(M) Un meganúmero es el número natural tal que no existe ninguno mayor que él.

Si aquí entendemos «mayor» en sentido cuantitativo, resulta que (M) es incoherente. No existe, en efecto, un número natural máximo, puesto que la serie de números naturales es infinita.

¿Hay alguna razón para suponer que la premisa (1) adolece de una incoherencia semejante? La habría en caso de que las cualidades que hacen grande a Dios no fueran máximas (es decir, fueran cualidades que pudieran ser poseídas siempre en un grado mayor, como sucede con la altura o el peso). Pero es plausible la suposición de que las cualidades que hacen grande a Dios son máximas. En efecto, las cualidades de omnipotencia, omnisciencia y bondad perfecta parece ser máximas; ningún ser puede ser más poderoso que un ser omnipotente, por ejemplo. Por consiguiente, no podemos criticar la definición que san Anselmo da de «Dios» de la misma manera en que criticamos la definición (M).

Se podría objetar, no obstante, que incluso si la premisa (1) fuera coherente, no podría tener consecuencias ontológicas. Una mera estipulación no puede generar entidades reales. Cuando la palabra «soltera» fue introducida por primera vez en nuestro idioma, y definida como equivalente a «mujer no casada», esa definición no garantizaba que el mundo contuviera solteras. El que haya solteras se debe no a definición alguna, sino al hecho (no verbal) de que algunas mujeres prefieren no casarse. De igual manera, el que tengamos definiciones significativas de palabras como «dragón» o «unicornio» no debería inducir a nadie a pensar que existen tales criaturas. Por lo tanto, ¿cómo podría la premisa (1), una mera definición, tener consecuencias ontológicas?

Sin embargo, insistir en este momento en esa objeción sería como dar por sentado que san Anselmo se equivoca, en lugar de demostrarlo. Pues Anselmo podría razonablemente replicar que, aunque sea cierto que muchas definiciones no tienen, en efecto, consecuencias ontológicas, esta definición suya sí las tiene. Para criticar ecuánimemente a san Anselmo, debemos examinar los pasos siguientes de su razonamiento.

En cuanto lo hacemos, surgen, sin embargo, las dudas. Enseguida salta a la luz que san Anselmo tiene una manera muy extraña de entender lo que sig-

nifica tener un concepto o una idea. Los tres primeros enunciados de la cita de arriba sugieren la siguiente concatenación de pensamientos. Primero entiendo una palabra «F» (digamos, un término general). En virtud de haber entendido «F», un F existe en mi entendimiento, que tiene todas las cualidades convencionales de un F. Y entonces podemos inquirir si hay instancias de F en la realidad.

Así pues, si entiendo el término «unicornio», habría un unicornio que existe en mi entendimiento, y ese unicornio tendría las cualidades que normalmente son asociadas a los unicornios (tener cuatro patas, un cuerno en forma de espiral, un rabo de león, etc.), aunque carecería de la cualidad de la existencia real. Pero esto es inverosímil. Cuando entiendo la palabra «unicornio», ¡no se trata de que yo tenga en mi mente algo de cuatro patas y con un cuerno en forma de espiral! Anselmo comete lo que pudiéramos denominar una *falacia de reificación*. Ha identificado el entender una palabra o aprehender un concepto con el que la mente contenga al objeto que está siendo concebido. Pero esto equivale a confundir un concepto con un objeto: el concepto está en mi mente, pero el objeto correspondiente no lo está. Los principios que subyacen a las premisas (2) y (3) son ejemplos patentes de la mencionada falacia. Una vez que los reconocemos como falacias —es decir, una vez que reconocemos que entender la palabra «Dios» no implica que haya en nuestra mente algo omnipotente, omnisciente, etc.—, no podemos por menos de rechazar aquellas dos premisas. Se basan en un modelo insostenible de lo que significa entender una palabra.

Aunque lo anterior basta por sí solo para anular la versión de san Anselmo del argumento ontológico, vale la pena citar la respuesta que le dio a éste uno de sus contemporáneos, Gaunilo de Marmoutiers. En su escrito titulado *En defensa del necio,* Gaunilo propone la idea de una isla tal que ninguna isla más excelente que ella puede ser concebida, y escribe luego:

> Ahora bien, si alguien me hablara de tal isla, entendería fácilmente sus palabras [...] Pero supongamos que esa persona realiza a continuación la siguiente especie de inferencia lógica: «No puedes seguir dudando de que dicha isla, que es más excelente que cualquier otra, exista en alguna parte, ya que no dudas de que está en tu entendimiento. Y puesto que es más excelente el no hallarse solamente en el entendimiento, sino el hallarse tanto en el entendimiento como en la realidad, por esta razón debe existir también. Pues si no existiese, cualquier isla que realmente existiera sería más excelente que ella; y por lo tanto, la isla que ya concebiste como la más excelente no sería la más excelente»[2].

Gaunilo intenta parodiar aquí la prueba de Anselmo. Es decir, está empleando un razonamiento análogo al de Anselmo para probar una conclusión obviamente absurda. Está claro que el mundo no contiene un isla perfecta (es decir, una isla tal que no quepa concebir una más perfecta que ella), o un caballo perfecto, o un cocodrilo perfecto, etc. Si existen argumentos análogos al de san Anselmo que son inválidos, también el argumento de Anselmo deberá ser

inválido. Nótese que los argumentos paródicos, aunque pueden ser eficaces en cierto sentido, son deficientes en otro. Cuando tienen éxito, un argumento paródico muestra que el argumento inicial (parodiado) está equivocado, pero no ofrece diagnóstico alguno acerca del origen del error.

Sin embargo, parece que la parodia de Gaunilo fracasa, porque su argumento no es análogo al de Anselmo en un aspecto fundamental. Mencioné antes que para que la definición de Dios dada por Anselmo fuera coherente, las cualidades que hacen grande a Dios deberían ser máximas, esto es, cualidades que no pueden ser poseídas en un grado mayor. Pero las cualidades que hacen que una isla sea excelente (tales como la abundancia de cocos, el número de palmeras, la presencia de bellas jóvenes de piel bronceada, etc.) no son, obviamente, cualidades máximas. Pero entonces la descripción «isla tal que ninguna más excelente pueda ser concebida» (así como la de «número natural tal que ninguno mayor pueda ser concebido») no expresa un concepto coherente.

### El argumento ontológico

El argumento ontológico de san Anselmo es un ejemplo clásico de un argumento racionalista. Intenta mostrar que podemos establecer una conclusión sustancial —la existencia de Dios— valiéndonos únicamente de la razón. Esto contradice el principio empírico (asociado a filósofos británicos como Locke, Berkeley y Hume) según el cual la razón nunca puede producir por sí sola un conocimiento sustancial de la realidad. Los empiristas sostuvieron que el conocimiento de la realidad se basa esencialmente en los datos de los sentidos. El argumento ontológico es ingenioso. Trata de probar la existencia de Dios recurriendo únicamente a la idea o definición de Dios como algo tal «que no se puede concebir nada mayor que ello». Sería un prodigio que la definición de una palabra pudiera probar la existencia de algo distinto de ella misma. Afortunadamente, el argumento de Anselmo no logra hacer tal cosa.

## El argumento cosmológico

Aunque en desacuerdo con el argumento ontológico de Anselmo, santo Tomás de Aquino (*ca.* 1225-1274) trató de probar la existencia de Dios por medio del llamado argumento cosmológico. Este argumento puede asumir muchas formas, una de las cuales es conocida como «el argumento de la contingencia». El filósofo e historiador jesuita Frederick Copleston, en un famoso debate con Bertrand Russell, delineó una versión del argumento de la contingencia:

el mundo es simplemente [...] el agregado de objetos individuales, ninguno de los cuales contiene en sí mismo la razón de su existencia [...] Puesto que ningún objeto de experiencia contiene en sí mismo la razón de su propia existencia, esa totalidad de objetos debe tener una razón externa a sí misma. Esta razón debe consistir en un ser existente. Ahora bien, o este ser es la razón de su propia existencia, o no lo es. Si lo es, estupendo. Si no lo es, debemos proceder igual que antes. Pero si procedemos hasta el infinito, [...] no habrá explicación alguna de la existencia. Por lo tanto, [...] para poder explicar la existencia, debemos llegar hasta un ser que contenga en sí mismo la razón de su propia existencia [...] un ser que no pueda no existir[3].

Más adelante en el debate, Copleston hace una recapitulación sucinta de la esencia del argumento:

La serie de eventos [incluyendo la historia del universo] está causada, o no lo está. Si lo está, es obvio que debe haber una causa fuera de la serie. Si no lo está, entonces se basta a sí misma; y si se basta a sí misma, esto es lo que yo denomino lo necesario. Pero aquél [el universo] no puede ser necesario, ya que cada uno de sus miembros es contingente, y [...] la totalidad no tiene realidad alguna aparte de sus miembros[4].

Para evaluar estas propuestas quizás sea útil comenzar por algunos de los comentarios críticos de Russell. Ambos argumentos se valen de nociones modales. La conclusión del razonamiento de Copleston es que existe un ser necesario (a saber, Dios). Russell discrepa en este punto: «La palabra "necesario" [...] sólo puede ser aplicada significativamente a proposiciones. Y [...] sólo a aquellas que son analíticas, es decir, las que resultan autocontradictorias cuando son negadas»[5]. Para Russell, ninguna proposición del tipo «A existe» puede ser analítica. Por lo tanto, hablar de «seres necesarios», o incluso de «seres contingentes», es un solecismo.

## Santo Tomás de Aquino (*ca*. 1225-1274)

Santo Tomás de Aquino nació en Roccasecca, Italia. Estudió las obras de Aristóteles en la Universidad de Nápoles y luego se convirtió en monje dominico. La obra más conocida de Aquino es su *Suma teológica* (1266-1273). Sostuvo que la fe y la razón no pueden discrepar entre sí, ya que la razón, cuando es ejercida adecuadamente, nunca dará resultados que contradigan la fe. Santo Tomás negó que la existencia de Dios pudiera ser probada mediante una mera reflexión sobre la idea de Dios; por lo tanto, rechazó el argumento ontológico de san Anselmo. Sin embargo, afirmó que la existencia de Dios podía ser establecida a partir de premisas relativas a la naturaleza y existencia del universo (*Suma teológica*, 1a, qu. 2, art. 3). Se trata de las famosas «cinco vías» para probar la existencia de Dios, una de las cuales se refiere a la existencia contingente del universo (que es una de las versiones del argumento cosmológico).

Por supuesto, si el argumento ontológico fuera válido, entonces «Dios existe» sería una proposición analítica; pero ya se vio que dicho argumento era defectuoso. Empero, incluso si concedemos que «Dios existe» no es una proposición analítica, sigue habiendo una réplica a Russell. Desde los trabajos de Kripke en las décadas de 1960 y 1970, se acepta generalmente que la analiticidad no es la única fuente de necesidad: habría verdades necesarias que no son analíticas[6]. Ya antes de Kripke, algunos filósofos habían sospechado que verdades tales como «nada puede ser rojo y verde en toda su superficie» eran necesarias pero no analíticas. Sin embargo, Kripke elaboró un marco teórico para explicar las necesidades no analíticas y proporcionó ejemplos más precisos, como por ejemplo «Tulio es Cicerón», «el agua es $H_2O$», «este atril es de madera», etc. Si esta pieza de ortodoxia contemporánea está en lo cierto, la no analiticidad de «Dios existe» no sería, por lo tanto, una razón válida para descartar su carácter necesario.

Pero aun hay un problema más grave en la respuesta de Russell. Es verdadero afirmar, con respecto a cualquiera de las cosas ordinarias con que nos solemos topar, «esto podría no haber existido». Incluso cabe decir de sí mismo: yo habría podido no existir (por ejemplo, si mis padres no se hubiesen conocido). Pero entonces se sigue inmediatamente que «yo existo» es una verdad contingente, y que, por consiguiente, yo soy un ser contingente. Así es que tendría sentido, después de todo, hablar de seres contingentes, en cuyo caso también lo tendría hablar de seres necesarios (independientemente de si hay o no tales seres).

Ocurre, sin embargo, que esta disputa no toca el meollo del asunto. Aunque Copleston usa, de hecho, expresiones tales como «ser... que no puede no existir» y «necesario» para describir a Dios, el aspecto de Dios relevante para su argumentación no es la necesidad sino la autosuficiencia: Dios contendría en sí mismo la razón de su propia existencia.

El problema es que la noción de algo que contiene la razón de su propia existencia es demasiado oscura como para que pueda servirnos de ayuda. Russell afirma que, a menos que «razón suficiente» signifique «causa», está ignoro de su significado. Pero «razón suficiente» no puede significar «causa», ya que, si así fuera, se tendría que poder decir que Dios es «causa de sí mismo»: algo que, según el propio Copleston admite, carece de sentido. Así pues, el reclamo de Russell parece justificado: si «razón suficiente» no significa «causa», ¿qué significa? Por eso, habría que aclarar esta noción para que pudiéramos evaluar la primera de las formulaciones de Copleston del argumento cosmológico.

En cuanto a la segunda de sus formulaciones, más escueta, ella no apela a la noción de autosuficiencia sino en un estadio avanzado del argumento. Comienza con la premisa: el universo (entero) posee o no posee una causa. Russell objeta que esta premisa carece de sentido. Tiene sentido, alega, preguntar por la causa de este o aquel evento (es así, en definitiva, como aprendemos el significado de la palabra «causa»); pero no tiene sentido preguntar por la causa de todos los eventos, la causa del universo entero.

Pero la razón que Russell esgrime a favor de su aserto no es convincente. Señala que es una falacia inferir que una totalidad posea determinada propiedad sólo porque cada miembro de ella la posee. Y existe, en efecto, una falacia parecida: la falacia de composición[7]. Para usar el ejemplo de Russell: del hecho de que cada ser humano tenga una madre no podemos inferir válidamente que la especie humana como tal tenga una madre. Pero sucede que Copleston no comete la falacia que le atribuye Russell. Su modo de razonar no es: todo evento tiene una causa; por ende, la totalidad de los eventos tiene una causa. Tan sólo piensa que tiene sentido preguntar si el universo tiene una causa. Y preguntar si el universo tiene una causa no parece involucrar el mismo tipo de error categorial que se comete al preguntar si la especie humana tiene una madre.

Por lo tanto, concedámosle a Copleston su premisa inicial: el universo entero tiene una causa o no la tiene. ¿Cómo prosigue el razonamiento? Dado que la noción de autosuficiencia resultó ser demasiado oscura, tenemos que operar con la noción de necesidad. Así pues, podemos refundir el segundo razonamiento de Copleston en los términos siguientes:

(1) El universo tiene una causa o no la tiene.
(2) Si el universo tiene una causa, la causa yace fuera de él.
(3) Si el universo carece de causa, el universo es necesario.
(4) El universo no es necesario.
(5) El universo tiene una causa (de [3] y [4]).
(6) Por lo tanto, la causa del universo yace fuera de él (por [2] y [5])[8].

Ahora que ya no tenemos reparos a la terminología empleada aquí, ¿cómo deberíamos responder? Las premisas (1) y (2) no presentan problemas: (1) es un truismo (ya que hemos admitido, *contra* Russell, que tiene sentido hablar de la posibilidad de que el universo tenga una causa); y (2) parece correcta, ya que si la causa del universo yaciera en él mismo, sería parte del universo. Pero ninguna parte de algo puede ser la causa de ese algo, por lo que la causa del universo debe yacer fuera del universo.

La premisa (4), aunque ha sido negada por algunos filósofos, también es plausible[9]. Si se admite que podría no haber existido nada, en lugar de existir algo, habría que admitir también que la existencia del universo es un asunto no necesario, es decir, contingente. Puede que la razón de Copleston para (4) tenga visos de incurrir en la falacia de composición: pues arguye que el universo entero es contingente porque cada uno de sus elementos es contingente. Pero su razonamiento no parece ser falaz en este caso. ¿Cómo podría una totalidad ser necesaria si cada uno de sus elementos es contingente? La falacia de composición bien podría ser una falacia sólo con respecto a ciertas clases de propiedad, entre las cuales podría no hallarse la propiedad de la contingencia.

Por lo tanto, si el argumento tiene algún fallo, éste tendrá que residir en la premisa (3). ¿Se puede sostener que el universo no tiene causa alguna y, sin

embargo, es contingente? *Prima facie,* sí. Uno puede afirmar, como lo hizo Aristóteles, que el universo siempre ha existido, pero que es contingente[10]. En segundo lugar, muchos cosmólogos contemporáneos creen que el universo, aunque no siempre ha existido, se originó a partir de la nada: no tendría, pues, causa alguna. Según esta concepción, el universo no tuvo una causa, y, sin embargo, su generación no fue necesaria. Un defensor del argumento cosmológico debería mostrar que concepciones como ésta son insostenibles; y mientras no lo haga, el argumento cosmológico nunca resultará del todo convincente. Además, aunque el argumento cosmológico fuera convincente, su conclusión no implicaría nada acerca de la naturaleza de la causa del universo (es decir, acerca de si esta causa es buena o amorosa).

## El argumento teleológico

El argumento teleológico a favor de la existencia de Dios, también denominado «argumento del designio divino», trata de argumentar a favor de la existencia de Dios partiendo no simplemente de la existencia contingente del universo, sino de lo intrincado y complejo de la estructura de éste.

Aunque el argumento ha estado circulando desde tiempos de los estoicos, William Paley (1743-1805) fue quien formuló la mejor versión de él, basada en la analogía siguiente. Paley alegó que cuando hallamos una piedra en el desierto, no se nos ocurre pensar que alguien la haya diseñado. Pero cuando nos topamos con un instrumento complejo como un reloj, suponemos que alguien lo fabricó, y estamos justificados en nuestra suposición. Pues al examinar sus partes, podemos percibir que están «elaboradas y dispuestas con un propósito determinado»[11]. Por consiguiente, podemos concluir que «el reloj ha debido de tener un fabricante... Ha tenido que haber, en algún lugar determinado, uno o varios artífices que lo configuraran con el propósito que nos consta que cumple actualmente; personas que comprendieran su mecanismo y previeran su uso»[12].

Paley argumenta a renglón seguido que podemos inferir análogamente, a partir del universo que encontramos a nuestro alrededor, y especialmente a partir de sus instancias biológicas particulares, tales como flores, corazones y ojos, la existencia de un diseñador del universo: Dios. Si estamos justificados en pasar de la existencia del reloj a la existencia del relojero, entonces nos asiste el mismo derecho a pasar de la existencia de estructuras biológicas complejas a la existencia de un creador. Es más, la belleza y la simetría de tales estructuras apuntarían hacia la naturaleza benevolente de su creador.

¿Cómo deberíamos responder a este argumento? Nótese que, a diferencia de los argumentos anteriores, el argumento teleológico no es presentado como si fuera deductivamente válido. Debido a que no es lógicamente imposible ni una creación *ex nihilo* ni un reagrupamiento azaroso de partículas en una configuración parecida a un reloj, la existencia de un reloj no implica lógicamen-

te la existencia de un relojero. La tesis de Paley es, más bien, que si uno se topa con un reloj, es razonable creer que fue construido por un relojero. Ésa sería la mejor explicación posible de la existencia del reloj. Análogamente, cuando vemos el mundo complejo que nos rodea, resulta razonable creer que fue creado por Dios.

### William Paley (1743-1805)

William Paley estudió y enseñó en la Universidad de Cambridge, antes de convertirse en clérigo de la Iglesia de Inglaterra. Escribió tres obras mayores. En sus *Principios de filosofía moral y política* (1785) defendió con razones teológicas el utilitarismo (la doctrina según la cual debemos hacer cuanto maximice la felicidad humana). En *Una concepción de la evidencia del cristianismo* (1794) disertó sobre la creencia en los milagros e hizo comentarios a la discusión sobre el tema en la sección 10 de la *Investigación sobre el entendimiento humano*, de Hume. La *Teología natural* (1802) fue la obra más conocida de Paley, y en ella presentó su versión del argumento de la providencia, alegando que la hipótesis de la creación divina es la mejor manera de explicar el mundo complejo y variado que nos rodea.

Lo primero que hay que observar es que incluso si se piensa que el razonamiento de Paley es persuasivo, lo más que tal razonamiento puede establecer es que Dios creó el universo, no que Dios aún existe. Es totalmente compatible con el argumento de Paley la suposición de que «Dios abandonó su obra» poco después del acto de la creación. Por ello, su argumento no nos proporciona razón alguna para pensar que Dios existe ahora, así como tampoco tenemos razón para pensar que el relojero siga existiendo ahora. (Curiosamente, Newton pensó que, sin Dios, los planetas saldrían despedidos en todas direcciones. Según este físico, el movimiento regular de los planetas hacía razonable creer en la existencia continua de Dios.)

En segundo lugar, en los *Diálogos sobre la religión natural* (1779), de David Hume, publicados tras la muerte de éste y antes de que apareciera, en 1802, la obra de Paley, uno de los interlocutores, Filón, señala cuán poco justificados estamos en extraer conclusiones acerca del creador a partir de nuestras observaciones sobre el (poco benigno) mundo que nos rodea:

> Nuestro mundo, como este individuo [el defensor del argumento del designio] debería saber, es muy deficiente e imperfecto, si lo comparamos con un ideal superior; y seguramente fue el primer ensayo de alguna deidad infantil, que luego lo abandonó a su suerte, avergonzado de su pobre desempeño; o es obra de un dios menor y subordinado, y motivo de burla de sus superiores jerárquicos; o es producto de la edad avanzada, herencia de alguna deidad senil, por lo que, fallecida esta última, ha estado dando tumbos de manera más o menos azarosa, desde que recibiera de aquélla su primer impulso y energía[13].

En efecto, como sugiere Filón, el argumento de la providencia no nos da razón alguna para suponer que hubo un solo creador. ¿Por qué no creer más bien que así como un «gran número de hombres se reúnen para construir una casa o un barco, edificar una ciudad o crear una alianza; [...] un cierto número de deidades se agrupan para diseñar y construir un mundo?»[14].

En tercer lugar, y lo que es más importante, debemos recordar que Paley estaba escribiendo antes de que apareciesen la teoría evolucionista de Darwin y la cosmología contemporánea. Para una persona de 1800, la única explicación posible del orden del universo era la creación divina. El razonamiento de Paley parecía entonces proceder del más elemental sentido común. En cambio, desde nuestro privilegiado punto de vista actual, disponemos de otra explicación —la teoría evolucionista de Darwin—, y el razonamiento de Paley ya no nos parece ineludible.

Cuando nos preguntamos cómo los seres humanos, con todos sus intrincados órganos (como los ojos y el corazón), se originaron, disponemos ahora de dos explicaciones alternativas: la teoría evolucionista de Darwin de la selección natural y la de la providencia divina. La teoría de la evolución ha obtenido un gran éxito en la explicación de los fenómenos, y está tan consolidada que nadie la pone seriamente en duda. Nos provee de la mejor explicación disponible (actualmente) de la existencia y naturaleza de las especies vivas. Ello determina el fracaso del razonamiento de Paley: el ojo no sería, después de todo, análogo a un reloj que nos encontramos en el camino. La presencia de la explicación darwiniana sobre el origen de las especies ha desbaratado completamente el argumento teleológico.

## El argumento del mal

Hasta este momento hemos venido criticando algunos argumentos a favor de la existencia de Dios. Ahora nos proponemos considerar un argumento muy conocido en contra de su existencia: el argumento del mal. Se trata de un argumento en contra de la existencia de Dios que presupone que Dios (en caso de que exista) posee los rasgos de omnipotencia, de omnisciencia y de bondad perfecta.

La idea básica que subyace al argumento del mal es que la existencia del mal, tanto si se trata del mal natural (como la muerte y el sufrimiento causado por los desastres naturales) como si se trata del mal moral (la crueldad mutua de los seres humanos), es incompatible con la existencia de Dios. Dado que el mal existe, no existiría Dios.

Sin embargo, tenemos que desmenuzar un poco las cosas para poder percibir con claridad la naturaleza de la supuesta incompatibilidad. El argumento del mal sostiene que las proposiciones siguientes forman un cuarteto incoherente:

(1) Dios es omnisciente.
(2) Dios es omnipotente.
(3) Dios es totalmente bueno.
(4) El mal existe.

Si (1)-(4) son incompatibles entre sí, no pueden ser todas verdaderas. Como (4) es sin lugar a dudas verdadera, una de las proposiciones (1)-(3) ha de ser falsa. Pero si cualquiera de (1)-(3) es falsa, entonces Dios no existe. Si no hay un ser que sea omnisciente, omnipotente y totalmente bueno, entonces Dios no existe. De acuerdo, pues, con el argumento del mal, no existe un ser tal y, por lo tanto, no existe Dios.

Pero ¿por qué pensar que (1)-(4) son incompatibles? Dijimos que el mal, tanto natural como moral, existe. Pero si Dios es omnisciente, sabe que el mal existe; si es totalmente bueno, quiere evitar el mal; y si es omnipotente, es capaz de evitarlo. Entonces, ¿por qué no lo evita? Si hubiera un ser que fuera omnisciente, omnipotente y completamente bueno, el mundo no contendría mal alguno. Pero el mundo contiene el mal. Por lo tanto, no existe un ser con esas características.

Este argumento posee una claridad encomiable, a pesar de que ha recibido, por supuesto, muchas réplicas. Éstas intentan, de modos diferentes, mostrar cómo (1)-(4) resultarían, después de todo, compatibles. He aquí una réplica posible. ¿Qué pasaría si algunos males fueran *necesarios* para que existieran ciertos bienes, o, dicho de otra manera, si algunos bienes no pudieran existir sin ciertos males? Entonces el creyente podría argumentar que el valor de los bienes en cuestión sobrepasa el costo de los males concomitantes. En tal caso, la existencia de los males no sería incompatible con el amor, el conocimiento o el poder divinos.

Un ejemplo sencillo ilustra esta idea: cuando estoy sentado en el sillón del dentista experimento dolor, pero se trata de un dolor necesario para poder tener dientes sanos y libres de dolor en el futuro. Sin embargo, esta necesidad —el hecho de que yo *deba* experimentar dolor ahora para poder verme libre de él en el futuro— es un caso de necesidad causal o tecnológica. En el estadio actual de la tecnología, las visitas dolorosas al dentista son indispensables para poseer dientes sanos en el futuro. Pero Dios es omnipotente. No está constreñido por necesidades causales o tecnológicas. La omnipotencia divina significa que Dios puede hacer todo cuanto sea lógicamente posible. Habría podido disponer las cosas de tal modo que las visitas al dentista fueran indoloras, o nos habría podido equipar con dientes incorruptibles. Entonces, ¿por qué no lo hizo?

Lo que la defensa del «mal necesario» requeriría, para tener visos de credibilidad, son males que fueran **lógicamente necesarios** para la existencia de ciertos bienes. Entonces, si fuera necesario que el mundo contuviera tales bienes, Dios no podría ser criticado por permitir los correspondientes males, ya que sería lógicamente imposible tener tales bienes sin los

concomitantes males. Ni siquiera Dios es capaz de hacer lo lógicamente imposible.

Ejemplos de bienes que parecen exigir lógicamente ciertos males son el heroísmo, la benevolencia y la compasión. El heroísmo, la benevolencia y la simpatía son posibles únicamente porque el mundo contiene, por ejemplo, desastres naturales, hacia cuyas víctimas otras personas pueden hacer gala de heroísmo, de compasión o de benevolencia.

Sin embargo, podemos hacer tres reparos al argumento anterior. En primer lugar, bienes como la compasión no parecen exigir la presencia de desastres reales. Sería suficiente con que se diera la apariencia de un desastre para suscitar el sentimiento de compasión. Si Dios fuera omnipotente, no le habría sido difícil crear un mundo en el que hubiera maremotos ficticios a los que pudiéramos reaccionar. Pero no lo hizo. ¿Por qué no?

En segundo lugar, la defensa del «mal necesario» es demasiado limitada. El mundo contiene muchos males que no poseen un bien correlativo o atenuante. Muchos casos de muerte y de sufrimiento a lo largo de la historia del hombre (y de los animales) ni siquiera han sido observados o registrados. Tales males no pueden, obviamente, ser justificados en términos de la producción de bienes en otras personas[15].

Tercero, hay algo moralmente cuestionable en la defensa del «mal necesario». ¿No están siendo usadas las víctimas de los desastres naturales como meros medios para el perfeccionamiento moral de los demás? «El sufrimiento ajeno es bueno para mi alma» difícilmente es aceptable como la consigna de alguien honesto.

Una segunda estrategia de réplica, destinada a mostrar la compatibilidad de (1)-(4), es la que se conoce como «defensa por recurso al libre albedrío». Dios dotó a los hombres de libertad de acción, y eso es un bien. Un mundo que contiene gente libre que a veces escoge libremente el bien y a veces escoge libremente el mal es preferible, se alega, a un mundo de autómatas que están programados para hacer siempre el bien. Pero la libertad tiene un precio: muchos males asolan al mundo como consecuencia de la libertad de acción. Puesto que Dios no es responsable de tales males, éstos no contradicen su bondad, su amor o su sabiduría.

Esta respuesta es una variante de la primera: para que pueda existir cierto bien (el libre albedrío) es necesario hacerse cargo de las malas consecuencias provenientes del mal empleo de ese bien. Sin embargo, la defensa que apela al libre albedrío está sujeta a las cinco objeciones siguientes:

(1) La defensa sólo abarca los casos de mal moral (el mal que resulta de una elección humana libre). No contribuye en nada a explicar por qué Dios admitiría la ocurrencia de males naturales.
(2) La defensa presupone que tenemos libre albedrío. Pero algunos filósofos han negado esto último, o bien porque son fatalistas, o bien por-

que piensan que el determinismo es verdadero o bien porque creen que el concepto de libre albedrío es confuso (véase el capítulo 7).

(3) No es obvio que el bien proveniente del libre albedrío compense la cantidad de mal que sabemos que los seres humanos han producido. Es poco plausible pensar que el valor de la libertad de una persona sobrepasa o compensa cualquier cantidad de mal producido por el ejercicio de ésta. ¿Acaso no sería mejor el mundo si Hitler o Stalin, por ejemplo, hubieran sido autómatas programados para hacer el bien?

(4) No es cierto que para que los hombres puedan tener libertad tiene que haber malas consecuencias. Incluso si Dios nos concedió el libre albedrío, habría podido intervenir cada vez que alguien hizo una elección mala, y asegurarse de que ésta no tuviera malas consecuencias. Eso habría sido sin duda mucho mejor que dejar que estas últimas ocurriesen. Entonces, ¿por qué Dios no intervino?

(5) Hay otra forma, más radical, en que Dios nos habría podido conceder el libre albedrío y garantizar que éste no tuviera malas consecuencias. ¿Por qué Dios no creó seres que siempre escogen libremente el bien?[16] Acaso se piense que tales seres son imposibles, pero eso es algo que dista de estar claro. ¿No escogen acaso los ángeles siempre el bien? Para no ir tan lejos, nosotros mismos somos seres que *a veces* escogemos libremente el bien. ¿Por qué habría de ser incoherente imaginarnos a nosotros mismos como si *siempre* eligiésemos el bien? Seguramente, no hay un límite superior a priori acerca del número de veces que un agente puede escoger libremente el bien.

## El problema del mal

La razón principal de por qué muchas personas sensatas no creen en Dios es la existencia del mal. No pueden comprender cómo un Dios omnisciente, omnipotente y supremamente bondadoso permitiría la proliferación del mal en el mundo (por ejemplo, el sufrimiento que sigue a un desastre natural como una inundación o una hambruna), o dejaría de impedir las consecuencias negativas de las malas decisiones de la gente. Los teístas (es decir, aquellos que creen en la existencia de Dios) han elaborado diversas teodiceas con la finalidad de demostrar cómo el conocimiento, el poder y el amor divinos son consistentes con la existencia del mal. Kant creyó que todas las teodiceas estaban condenadas al fracaso (*La religión dentro de los límites de la mera razón*, 1793), y nuestra discusión parece confirmar que estaba en lo cierto.

## Observaciones finales

Todos los argumentos convencionales a favor de la existencia de Dios están sujetos a objeciones. La existencia del mal proporciona un argumento contra la existencia de Dios. Por lo tanto, es bastante razonable suponer que no existe un ser que sea omnipotente, omnisciente y completamente bueno.

Podemos presentar esta última objeción del modo gráfico siguiente. Imaginemos que Dios pasara revista a todos los mundos posibles, preguntándose a cuál de ellos debería conceder realidad[17]. En algunos de esos mundos, todos eligen libremente el bien. En otros, nadie escoge el bien. Y en otros, algunas personas escogen el bien y otras escogen el mal. ¿Por qué Dios no eligió actualizar uno de los mundos de la primera clase? Si era capaz de hacerlo, quería hacerlo y sabía cómo hacerlo, lo habría hecho. Pero no lo hizo; por lo tanto, Dios carece de benevolencia, de conocimiento o de poder.

Una última estrategia para responder al argumento del mal señala que (1)-(4) no son formalmente inconsistentes. Quizás haya una razón por la que Dios permite el mal que nosotros no conozcamos ni podamos conocer. Sin embargo, esta respuesta no parece ser otra cosa que la expresión de un deseo.

## Preguntas para el estudio

- ¿Comete san Anselmo realmente la falacia de reificación al formular el argumento ontológico?
- ¿Tenía razón Russell en cuestionar la coherencia de la noción de ser necesario?
- ¿Puedes construir una versión más plausible del argumento del designio basándote en la enorme improbabilidad de un universo que permita la vida?
- ¿Cuál es la réplica más plausible que un teísta podría formular frente al argumento del mal?

### Notas

1. Anselmo, *Proslogion II,* en: *St Anselm's Proslogion,* M. Charlesworth (ed.) (Oxford: Oxford University Press) 1965, p. 116.
2. Gaunilo, «On Behalf of the Fool», en: *St. Anselm's Proslogion,* pp. 163-165.
3. «Must God Exist?», en *Philosophy in the Open* (Milton Keynes: Open University Press) 1978, p. 115.

4. Ibíd., pp. 119-120.
5. Ibíd., p. 115.
6. Véase S. Kripke, *Naming and Necessity* (Oxford: Basil Blackwell) 1980.
7. Esta falacia debe ser distinguida de la falacia de división, esto es, la que consiste en inferir que cada miembro de un todo tiene una propiedad por el hecho de que la totalidad la tenga.
8. La conclusión de que la causa del universo yace fuera de éste no implica por sí sola que la causa del universo sea un ser necesario. Sin embargo, los partidarios del argumento cosmológico normalmente presuponen que cualquier cosa que exista fuera del universo es un ser necesario.
9. Véase la discusión del realismo modal de David Lewis en el capítulo 2.
10. Esto no es una contradicción. El que algo siempre haya existido no implica nada sobre su correspondiente estatuto modal (es decir, sobre si es necesario o contingente). Es difícil encontrar objetos contingentes y eternos, pero la categoría como tal es coherente.
11. W. Paley, *Natural Theology,* ed. de F. Ferre (Nueva York: Bobbs-Merrill Co. Inc.) 1963, pp. 1-2.
12. Ibíd., p. 2.
13. D. Hume, *Dialogues Concerning Natural Religion,* ed. N. Kemp Smith (Londres: Nelson), p. 168.
14. Ibíd., p. 167.
15. Se ha sugerido a veces que a pesar de que un hombre sufra en la tierra, posee un alma inmortal que le permitirá disfrutar de una felicidad infinita en la vida después de la muerte. Pero ¿es esto realmente una réplica al argumento del mal? ¿Cómo justifica este argumento el sufrimiento terrenal?
16. Aquí estoy en deuda con el excelente artículo de J. L. Mackie titulado «Evil and Omnipotence», *Mind*, vol. 64, 1955, pp. 200-212.
17. El realista moral no aceptaría esta manera del plantear el asunto. (Véase el capítulo 2.)

## Bibliografía adicional comentada


M. M. Adams y R. M. Adams (eds.), *The Problem of Evil* (Oxford: Oxford University Press) 1992. Un interesante conjunto de ensayos, a ratos exigente, dedicado al problema del mal. La Introducción es muy útil, y yo en particular recomendaría los artículos de J. L. Mackie, N. Pike, A. Plantinga y R. M. Adams.

J. L. Mackie, *The Miracle of Theism* (Oxford: Clarendon Press) 1982. Una introducción excelente a todos los argumentos principales a favor y en contra de la existencia de Dios. Muy recomendable.

A. Plantinga (ed.), *The Ontological Argument* (Nueva York: Anchor Books) 1965. Un librito útil, dividido en dos partes. En la primera, Plantinga presenta una antología de distintas formulaciones del argumento ontológico, tomadas de san Anselmo, Descartes, Spinoza y Leibniz, junto con comentarios críticos de Gaunilo, Caterus, Gassendi, Kant y Schopenhauer. En la segunda, Plantinga hace una selección (hoy algo desactualizada, pero aún interesante) de comentaristas del siglo XX. La contribución más interesante es la de N. Malcolm, con réplicas de A. Plantinga y P. Helm.

A. Plantinga, *God, Freedom and Evil* (Londres: Allen & Unwin) 1975. Otro libro útil, también en dos partes. En la primera, Plantinga ofrece su propia teodicea ingeniosa, esto es, su solución al problema del mal, fundamentada en su versión de la defensa basada en el libre albedrío. En la segunda parte, esboza y critica las versiones estándar de los argumentos cosmológico, teleológico y ontológico a favor de la existencia de Dios. Termina proponiendo una versión del argumento ontológico que, en su opinión, asegura no la verdad, pero sí la aceptabilidad racional del teísmo. Ambas discusiones echan mano del concepto de mundo posible.

W. L. Rowe, *The Cosmological Argument* (Princeton, NJ: Princeton University Press) 1975. Una discusión muy clara, detallada e históricamente documentada de todos los aspectos del argumento cosmológico.

## Recursos en Internet

A. Plantinga (1998) «God, Arguments for the Existence of», en E. Craig (ed.), *Routledge Encyclopedia of Philosophy.* Bajado el 31 de mayo de 2006 de <http://www.rep.routledge.com/article/K029>.

G. Oppy, «Ontological Arguments», *The Stanford Encyclopedia of Philosophy (edición de otoño de 2005),* Edward N. Zalta (ed.). Bajado el 31 de mayo de 2006 de <http://plato.stanford.edu/archives/fall2005/entries/ontological-arguments>.

M. Tooley, «The Problem of Evil», *The Stanford Encyclopedia of Philosophy (edición del invierno de 2004),* Edward N. Zalta (ed.). Bajado el 31 de mayo de 2006 de <http://plato.stanford.edu/archives/win2004/entries/evil>.

# 2. La existencia

## Introducción

La existencia da lugar a una gran cantidad de interrogantes típicamente filosóficos que yacen en el corazón mismo de la metafísica. En este capítulo abordaremos los siguientes. En primer lugar, ¿cuál es la extensión o el ámbito del ser?[1] Y en especial, ¿existen las entidades no actuales o meramente posibles? ¿Es lo que existe de forma actual sólo una parte de todo lo que existe? En segundo lugar, ¿por qué existe el universo? ¿Por qué hay algo en lugar de nada? Y ¿por qué se da *este* algo en lugar de alguna otra cosa? En tercer lugar, ¿hay objetos no existentes? ¿Por qué alguien habría de pensar que los hay? En cuarto lugar, ¿qué es la existencia? Y, en especial, ¿es la existencia una propiedad de los objetos ordinarios o no? ¿Por qué es importante la forma en que contestemos esta pregunta? Responder los interrogantes 3 y 4 nos exigirá profundizar en cuestiones complejas referidas al significado, la denotación y la estructura lógica.

## El gran ámbito del ser

Es muy natural pensar que existe sólo lo que existe de manera actual. Por supuesto, podría haber cerdos voladores, dragones que arrojan llamas, montañas de oro, máquinas de movimiento perpetuo, y así sucesivamente. Pero tales cosas no existen. Habrían podido hacerlo, pero no lo hacen. David Lewis, sin embargo, no está de acuerdo. Lo que habría podido existir, dice, existe, en al-

gún otro mundo posible. Los mundos posibles existen también, exactamente de la misma manera en que lo hace el mundo actual. En eso reside la tan criticada doctrina del realismo **modal**. Lewis ha escrito:

> Abogo por una tesis de la pluralidad de los mundos o del *realismo modal,* la cual sostiene que nuestro mundo es un mundo entre otros. Hay incontables mundos alternativos, cosas de tipo muy incluyente [...] Están aislados: no hay en absoluto relaciones espaciotemporales entre cosas distintas que pertenecen a mundos diferentes. Y lo que sucede en un mundo no puede hacer que suceda una cosa en otro [...] Los otros mundos son de la misma especie que este mundo nuestro [...] Este mundo tampoco difiere de otros en su manera de existir [...] Los mundos no son productos nuestros[2].

La doctrina del realismo modal se puede sintetizar en estas siete tesis:

1. Los mundos posibles (incluyendo el nuestro) existen.
2. Son de la misma especie que nuestro mundo.
3. Los mundos posibles, al igual que el mundo actual, no pueden ser reducidos a otra cosa.
4. No hay nada ontológicamente especial respecto del mundo actual. Cada mundo es actual *para sus habitantes.*
5. «Actual» es simplemente una expresión deíctica, como «aquí».
6. Los mundos posibles están espaciotemporalmente aislados y, por consiguiente, causalmente aislados entre sí.
7. Los mundos posibles no dependen de la mente.

Es importante tener claras las implicaciones de la doctrina del realismo modal. Éste no es la concepción de que «muchos mundos existen actualmente» (una concepción a veces propuesta por algunos intérpretes de la mecánica cuántica). Según esta última, el mundo actual sería mucho mayor de lo que creemos que es, y eso no es lo que piensa Lewis. El realismo modal tampoco es la concepción de que hay mundos posibles o seres posibles que no existen. Para Lewis no hay diferencia entre «hay Fs» y «los Fs existen». Su tesis es que el conjunto de las cosas que actualmente existen es sólo un subconjunto de todas las cosas que existen, del mismo modo que el conjunto de las personas que viven en Camberra es un subconjunto de toda la población mundial, o el conjunto de las personas que estaban vivas en 1940 es un subconjunto de las personas que vivieron en el siglo XX.

Cada mundo posible corresponde a una forma diferente en que podría haber sido nuestro mundo. Algunos mundos posibles son como nuestro mundo: al igual que éste, contienen hombres y asnos de carne y hueso. Pero otros mundos posibles son más exóticos, y contienen tipos de entidades jamás imaginadas en este mundo. Los mundos posibles son mundos concretos como el nuestro, irreducibles a cualquier otra cosa (por ejemplo, a objetos abstractos tales como conjuntos de proposiciones).

De acuerdo con el realismo modal, el modo de existencia actual (o sea, el que es característico del mundo actual) no tiene estatuto privilegiado alguno. El modo de actualidad no es un indicio de lo real, ya que otros mundos son igualmente reales. La palabra «actual» proferida por un habitante de cualquier mundo se refiere simplemente al mundo en que vive ese habitante. Cada mundo es actual para sus habitantes, así como cada lugar es un «aquí» para quien lo ocupa. «Actual» es meramente una **expresión deíctica**.

Además, Lewis piensa que los mundos posibles son unidades espaciotemporales, y que éstas se hallan espaciotemporalmente aisladas unas de otras. No existe una distancia espacial o temporal entre los mundos posibles[3]. Finalmente, los mundos posibles no son ficticios, no son invención nuestra. Todos los mundos posibles (incluido el nuestro) existen independientemente de nosotros: no somos la causa de que se hayan originado ni de que sigan existiendo.

### El realismo modal

La doctrina del realismo modal procede en dos pasos. Primero analiza el habla modal ordinaria en términos del lenguaje del mundo posible. La oración verdadera «podría haber habido cisnes azules» es vertida como «hay un mundo posible que contiene cisnes azules». En segundo lugar, suscribe el realismo acerca de los mundos posibles y sus pobladores. Otros mundos posibles existen exactamente de la misma manera en que lo hace el nuestro. Los cisnes azules existen, pero no en nuestro mundo. Algunos filósofos niegan que el habla modal conlleve una cuantificación implícita sobre otros mundos. Otros aceptan la traducción al lenguaje de los mundos posibles, pero niegan que estos últimos existan. El mundo actual tiene un estatuto privilegiado: es el único que existe. Pero entonces, ¿qué haría que fuera verdadero que haya un mundo posible que contiene cisnes azules, si no fuera por la existencia de un mundo concreto que contiene cisnes azules? Los actualistas, como son conocidos, postulan, en lugar de un mundo, algún hacedor de verdad sucedáneo: por ejemplo, un conjunto completo y consistente de oraciones entre las que se encuentra la oración «Los cisnes son azules». Este conjunto sería lo que hace verdadero que hubiera podido haber cisnes azules. Pero ¿es esto plausible? Incluso si no hubiera habido semejante conjunto de oraciones (por ejemplo, si no hubieran existido mentes), habría podido haber cisnes azules.

## ¿Por qué existe nuestro mundo?

El realismo modal de Lewis tiene implicaciones para una de las preguntas más enigmáticas sobre la existencia: ¿Por qué existe nuestro mundo? Esta pregunta se puede dividir, por conveniencia, en dos interrogantes: ¿Por qué hay algo en vez de nada? y ¿Por qué hay *esto* en lugar de alguna otra cosa?

Lewis tiene una respuesta rápida al primer interrogante. Para Lewis no podría haber

> [...] un mundo completamente vacío. Un mundo no es como una botella vacía de cerveza. El mundo es la totalidad de las cosas que contiene [...] Si no existe ni siquiera la botella, entonces no habrá nada en absoluto [...] Podría no haber mucho [en un mundo]: quizás algún segmento espaciotemporal sin ocupar, o quizás sólo un punto dentro de él. Pero «no mucho» ya es algo, y no hay mundo alguno donde no haya absolutamente nada. Eso hace que sea necesario que haya algo[4].

Lewis no considera esto como una explicación de por qué hay algo en lugar de nada (él piensa que no existe tal explicación). Simplemente cree estar mostrando que la pregunta está basada en un falso supuesto: el de que hubiera podido no haber absolutamente nada, un mundo absolutamente vacío.

Sin embargo, incluso si Lewis se equivocase a este respecto y un mundo vacío pudiese existir, su doctrina del realismo modal seguiría teniendo implicaciones para la pregunta de por qué hay algo en lugar de nada. Si existe un mundo posible vacío, todavía podemos responder: el «en lugar de» pierde relevancia: habría la nada (en el mundo vacío posible). Según el realismo modal, ese mundo existiría tanto como el nuestro.

En cuanto a la pregunta «¿por qué *esto* en lugar de alguna otra cosa?», quien la formula tiene normalmente en mente una pregunta más específica como «¿por qué este mundo contiene la vida en lugar de no contenerla?». Derek Parfit ha descrito muy bien de dónde proviene nuestro interés por esta pregunta:

> Para hacer la vida posible, habría que seleccionar las condiciones iniciales con un tipo de precisión como el que se requiere para acertar en un blanco pequeño situado en una galaxia distante. Debido a que no es arrogante considerar la vida como algo muy especial, esta aparente afinación necesita ser explicada. De entre la infinidad de condiciones iniciales posibles, ¿por qué las que permiten la vida fueron *justamente* las que efectivamente se dieron?[5].

Una respuesta obvia es la teística: Dios, cuya existencia es necesaria, dispuso que las cosas fueran de esa manera. Pero evidentemente se trata de una buena respuesta sólo si existe Dios. (Véase al respecto el capítulo 1.) Otra respuesta es que no hay respuesta: sucede sencillamente que el universo actual permite la vida, y que eso no es algo que se pueda explicar. Pero esta respuesta no suele ser satisfactoria. Afortunadamente, hay una tercera opción:

> Nuestro universo podría no ser toda la realidad. Algunos físicos sugieren que hay muchos otros universos; o, para evitar confusiones, *mundos*. Estos mundos tienen las mismas leyes naturales que nuestro propio mundo, y emergieron de Big Bangs similares, pero cada uno tiene condiciones iniciales ligeramente diferentes. En esta *hipótesis de la abundancia de mundos* no habría necesidad de afinaciones. Si hubiera un núme-

ro suficiente de Big Bangs, no tendría nada de sorprendente que en algunos de ellos estuvieran dadas las condiciones para la vida. Y no sería sorprendente que nuestro Big Bang formara parte de ese pequeño grupo[6].

Aunque es posible que Parfit no tuviera en mente la doctrina del realismo modal de Lewis cuando escribió estas líneas, el realismo modal expresa la misma idea de otra manera. Si el realismo modal es verdadero, y nuestro mundo se cuenta como uno más entre un sinnúmero de mundos, entonces no es tan sorprendente que éste albergue la vida. Muchos otros lo harían, y el nuestro sólo sería uno de ellos. Por lo tanto, el realismo modal haría innecesaria una explicación teística. Dado que el realismo modal explica algo que de otra manera resultaría enigmático, tiene al menos esta ventaja a su favor.

### ¿Por qué hay algo en lugar de nada?

Durante siglos, muchos han pensado que esta pregunta exige una respuesta, y que sólo puede ser contestada recurriendo a Dios. Habría algo en lugar de nada porque Dios decidió crear este universo y las criaturas que éste contiene. Pero incluso si creemos que no existe Dios, la pregunta parece razonable. Hay, por supuesto, algunos que no la consideran razonable, y piensan que no se le puede dar una respuesta interesante. Coincidieron circunstancias iniciales que permitieron la emergencia del universo físico y la evolución de la vida. Pero no habría una explicación de por qué fueron justamente esas circunstancias iniciales las que se dieron. Muchos se conforman con eso, pero otros no. Si la probabilidad de que se den circunstancias iniciales que permitan la vida es tan baja, hasta un grado de inverosimilitud, ¿no tendría que haber una explicación de por qué dichas circunstancias se dieron? Cuando se formula esta pregunta, se presupone que hay un solo mundo: el mundo actual. Pero si esta suposición fuera falsa, como creen los realistas modales, nuestro mundo no sería más que uno entre muchos mundos que permiten la vida. En ese caso, no sería tan misterioso que nuestro mundo permitiera la vida: muchos mundos lo harían, y el nuestro sería tan sólo uno más entre ellos.

## Examen del realismo modal

Lewis afirma que el realismo modal posee una serie de virtudes entre las que destaca la de permitir un análisis ingenioso de las expresiones modales que empleamos al hablar (expresiones como «podría», «debo», «posiblemente», «necesario», etc.) y de nuestro uso de los enunciados contrafácticos (o sea, condicionales de la forma «Si A no hubiese ocurrido, tampoco habría ocurrido B»). Así, todos entendemos el enunciado «Los cerdos no vuelan», y sabe-

mos qué es lo que lo hace verdadero: cómo son las cosas actualmente. Sin embargo, también entendemos el enunciado «los cerdos podrían volar», y consideramos que es verdadero. Pero ¿qué hace que lo sea? No, presumiblemente, el estado de cosas actual. ¿Qué, entonces? Lewis ofrece una respuesta: el modo de ser de las cosas en otro mundo posible (muy diferente del nuestro). El enunciado «los cerdos podrían volar» es reformulado por Lewis como «hay un mundo posible en el que los cerdos vuelan». Dado que existe ese mundo, y dado que ese mundo es un mundo concreto como el nuestro, el misterio acerca de qué es lo que hace que verdades modales como ésa sean verdaderas quedaría resuelto. De manera similar, enunciados de la forma «posiblemente P» son reformulados como «hay un mundo posible en el que P», y enunciados de la forma «necesariamente P» son reformulados como «en todos los mundos posibles, P». «Posiblemente» y «necesariamente» serían, por lo tanto, asumidos como **cuantificadores** que rigen mundos posibles.

Surge, sin embargo, una duda en el caso de aseveraciones modales sobre individuos particulares y actuales. Supongamos que digo «Gore habría podido ganar las elecciones presidenciales estadounidenses de 2000», y que eso es verdadero. ¿Qué es lo que lo hace verdadero? Al igual que antes, Lewis responde: la manera en que las cosas son en otro mundo posible. Pero ¿cómo son las cosas en ese otro mundo posible? ¿Se encuentra entre ellas el propio Gore, alegrándose de su victoria? Ésa sería una manera de ver la cuestión. Según la teoría de la existencia múltiple [*multiple-existence theory*], los individuos existirían en más de un mundo posible.

Sin embargo, Lewis no opina lo mismo. Para él, como para Leibniz, los individuos están ligados a un mundo. Gore existe en un mundo (el nuestro) y en ningún otro. ¿Cómo, entonces, podría un mundo posible que no contenga a Gore ser el hacedor de verdad [*truth-maker*] del enunciado «Gore habría podido ganar las elecciones estadounidenses de 2000»? La respuesta de Lewis es que ese otro mundo contendría un *correlato* de Gore (alguien muy similar a Gore, pero que no es él) que ganaría las elecciones de 2000. Por lo tanto, el que Gore habría podido ganar las elecciones presidenciales de 2000 es cierto porque hay un mundo posible, similar al nuestro, que contiene un correlato de Gore que gana las elecciones.

He ahí, pues, el contraste entre la teoría de la existencia múltiple y la teoría correlacionalista [*counterpart theory*]. Ha sido célebre la protesta de Saul Kripke de que la teoría correlacionalista no consigue hacer justicia a nuestras aseveraciones modales:

> [De acuerdo con la teoría correlacionalista,] si decimos «Humphrey habría podido ganar las elecciones...», no estamos hablando de algo que le habría podido pasar a *Humphrey*, sino a otra persona, a un «correlato». Es probable, sin embargo, que a Humphrey le traiga sin cuidado si alguna otra persona, por mucho que se le parezca, obtiene o no la victoria en otro mundo posible[7].

Se ha discutido prolijamente si la objeción de Kripke es o no acertada. Los teóricos correlacionalistas responden que el que un correlato de Humphrey gane las elecciones es justamente lo que significa que sea verdadero que Humphrey habría podido ganar las elecciones. Eso es algo que niega el teórico de la existencia múltiple. No está claro quién sale airoso de este debate. Pero podemos defender una tesis más débil. La teoría de la existencia múltiple proporciona una semántica más fluida de nuestro discurso modal que su rival teórico. Según la teoría de la existencia múltiple, «Gore habría podido ganar las elecciones» es verdadero sólo si hay un mundo posible en el que Gore gana las elecciones. Ésta es una explicación más sencilla que la del teórico correlacionalista, y constituye un punto a favor de la teoría de la existencia múltiple.

Muchos filósofos han criticado el realismo modal que ha estado sirviendo de telón de fondo a la discusión precedente. Debido a su profusión ontológica, el realismo modal es una doctrina que cuenta con pocos partidarios. Muchos lo encuentran enormemente inverosímil. Y no es de extrañar. Se nos pide que creamos que hay mundos que contienen asnos que hablan y cerdos voladores, y que estos últimos existen exactamente en el mismo sentido en que existen nuestros asnos y cerdos. Pero si respondemos así, deberíamos reconocer que Lewis nos plantea un desafío: el de que expliquemos qué es lo que hace que nuestras aseveraciones modales sean verdaderas, sin recurrir para ello a mundos posibles construidos de manera realista. Quizás no sea imposible salir airoso de él. Por ejemplo, según la concepción narrativa [*the story view*], un enunciado como «los cerdos podrían volar» es verdadero si el enunciado «los cerdos vuelan» es miembro de alguna narración completa y consistente (un conjunto de enunciados). Pero los defensores del realismo modal objetarán que las posibilidades son objetivas e independientes de la mente: incluso si no hubiera habido enunciados, habría cerdos voladores. Por lo tanto, añadirán, la concepción narrativa no puede ser correcta. El debate continúa.

¿Qué deberíamos concluir acerca del realismo modal de Lewis? Se trata de una doctrina con ventajas explicativas importantes. Proporciona hacedores de verdad inequívocos para nuestras aseveraciones modales. Socava un supuesto de la pregunta enigmática «¿por qué hay algo en lugar de nada?» y permite responder sin rodeos a la pregunta «¿por qué este mundo contiene la vida en lugar de carecer de ella?». Sin embargo, estas ventajas son contrarrestadas por su desmesura ontológica. Se espera que creamos que los dragones, los duendes y las montañas de oro existen exactamente de la misma manera en que lo hacen los caballos, los árboles y las rocas. Se trata de algo difícil de admitir. Sin embargo, quien rechace el realismo modal nos queda debiendo una explicación alternativa sobre los hacedores de verdad de las aseveraciones modales, y está obligado a decirnos por qué, de todos los mundos posibles, el nuestro habría de tener el privilegio de existir.

## Objetos no existentes

Puede ser que Lewis se aparte del sentido común cuando se refiere a lo que existe, pero al menos no afirma que haya cosas inexistentes. En cambio, Colin McGinn sí lo hace. Es un defensor contemporáneo de la idea de que hay objetos inexistentes. En un libro reciente ha escrito:

[...] nos parece natural hablar de la siguiente manera. No todo aquello a lo que nos referimos existe: Venus lo hace, pero Vulcano no; los caballos lo hacen, pero los unicornios no. Hay entidades meramente ficticias, y hay cosas que realmente existen. Existir es poseer una propiedad que sólo tienen algunas de las cosas a las que nos referimos: las que existen, a diferencia de las que son meramente ficticias[8].

Al proponer este punto de vista, McGinn está siguiendo los pasos del filósofo austríaco Alexius Meinong, quien no vaciló en defender lo inexistente, amén de muchas otras cosas. ¿Por qué Meinong sostuvo una opinión tan insólita? He aquí un famoso pasaje de Bertrand Russell:

Se ha argumentado —por ejemplo, por Meinong— que podemos hablar de la «montaña de oro», «el cuadrado redondo», etc.; que podemos formular proposiciones verdaderas que tienen estas expresiones como sujeto, y que, por lo tanto, éstas deben tener algún tipo de existencia lógica, ya que de otra manera las proposiciones en las que aparecen carecerían de significado. En tales teorías, me parece, falla ese sentido de la realidad que debería preservarse hasta en los estudios más abstractos. La lógica no debería mostrar más tolerancia hacia los unicornios que la zoología; pues la lógica se ocupa del mundo real no menos que lo hace la zoología, aunque empleando rasgos más abstractos y generales. Decir que los unicornios tienen una existencia en la heráldica, en la literatura o en la imaginación es una evasiva extremadamente lamentable y ridícula[9].

El propósito central de Russell en el capítulo del cual fue extraída esta cita, así como en su artículo, más famoso, de 1905 «On Denoting» [Sobre la denotación], es frustrar la coherencia del argumento de Meinong[10]. Del mero hecho de que un sujeto gramatical sea significativo y figure en oraciones verdaderas y significativas no se sigue que haya algo a lo que se refiere. Por lo tanto, Russell rechaza decididamente la idea de que el significado sea lo mismo que la referencia. Decidida, mas no completamente: Russell piensa que hay un rango restringido de términos singulares genuinos para los que el significado equivale a la referencia (por ejemplo, el pronombre de la primera persona del singular «yo» o aquellos términos que se refieren a las experiencias inmediatas de cada uno, tales como «este dolor de cabeza»). Para Russell, términos como éstos son los únicos términos denotativos[11]. El uso adecuado de ellos («nombres propios desde el punto de vista lógico», como se los conoce) garantiza que tengan un referente.

## Alexius Meinong (1853-1920)

Alexius Meinong fue un filósofo austriaco, muy influido por su maestro, Franz Brentano (1837-1917). A partir de 1889, Meinong enseñó en la Universidad de Graz, e hizo importantes contribuciones a la filosofía y a la psicología filosófica. La obra de Meinong parte de su filosofía de la mente y, en particular, de la tesis de Brentano sobre la intencionalidad, la tesis de que una característica de lo mental y de los estados mentales es estar dirigidos hacia objetos. Esta tesis indujo a Meinong a proponer una teoría a gran escala de los objetos que incluye objetos posibles (la montaña de oro), objetos imposibles (el triángulo cuadrado) y objetos incompletos (algo alto). Cualquier sujeto de una proposición verdadera es un objeto. Para Meinong, la oración «el triángulo cuadrado es un triángulo» es verdadera, y por lo tanto hay un triángulo cuadrado. Estos objetos son independientes de la mente, pero como objetos potenciales de pensamiento. Meinong también creyó en valores objetivos, tales como el bien y la belleza, que serían detectables por medio de las emociones y los deseos.

Todos los demás términos gramaticales susceptibles de cumplir la función de sujeto (los nombres propios ordinarios, las descripciones definidas, etc.) son impostores. Dan la apariencia de estar ejerciendo la función denotativa, pero no lo hacen en absoluto. Las oraciones que los contienen poseen una estructura gramatical que no coincide con su verdadera **estructura lógica**. Y la intención de la teoría russelliana de las descripciones es exhibir su verdadera estructura lógica.

En el caso de algunas descripciones definidas (como expresiones de la forma «el tal-y-tal»), es claro que la estructura gramatical y la estructura lógica difieren entre sí. Considérese:

(1) La familia media tiene 2,3 hijos.

Desde un punto de vista gramatical, tenemos aquí un sujeto y un predicado (de la forma «Fa», donde «a» es el término-sujeto y «F» es el predicado). Pero el término-sujeto «la familia media» es un término singular espurio. Su función no es denotar a una familia en particular de la que luego se dice que tiene 2,3 hijos. Cualquiera que piense eso habrá entendido mal el significado de esta oración. La estructura lógica de (1) es exhibida por:

(2) El número de hijos dividido por el número de familias = 2,3.

(1) es simplemente una forma abreviada de expresar (2), y (2) expresa el resultado de una división aritmética. Tiene la forma «a / b = c», y no «Fa». Adicionalmente, en (2) no hay un término destinado a referirse a la familia media, y eso, para Russell, es un signo inequívoco de que la descripción «la familia me-

dia», aunque sea gramaticalmente un término-sujeto, no es un término singular genuino. (Para Russell, la denotación genuina es irreducible.) Aunque este término-sujeto tenga un significado, su función no es la de denotar algo.

Eso basta para revelar un error en el razonamiento de Meinong. Pero Russell se proponía mostrar además que el razonamiento de Meinong adolece de un fallo generalizado, no limitado al caso de descripciones específicas como «la familia media». La mayoría de la gente entiende las convenciones que subyacen a la forma «el tal-y-tal promedio es F», y es poco probable que caiga en el engaño de creer que el término-sujeto contenido en ellas es un término denotativo. Pero hay casos que se prestan a confusión. Por lo tanto, Russell quería que su teoría mostrara que ninguna descripción (definida o indefinida) es un término denotativo genuino. La estructura gramatical de oraciones que contienen descripciones no sería en absoluto equiparable a su estructura lógica.

De acuerdo con la teoría russelliana de las descripciones, las descripciones son cuantificaciones existenciales encubiertas. Cuando las descripciones aparecen en oraciones de la forma «un F es G» y «el F es G», no son términos singulares genuinos, aunque constituyan términos-sujeto gramaticales. La verdadera estructura lógica de «un F es G» es captada por $\exists x$ (F$x$ y G$x$) (es decir, algo es a la vez F y G). La verdadera estructura lógica de «el F es G» está captada por $\exists x$ [F$x$ y ($y$) (si F$y$, entonces $x = y$) & G$x$] (es decir, hay un $x$ tal que es F, y $x$ es la única cosa que es F, y $x$ es G).

En los resultados de estos análisis no aparecen términos singulares denotativos; sólo la cuantificación (el cuantificador existencial «hay» y el cuantificador universal «para todos»), la predicación («F» y «G») y la identidad («=»). No hay nada en ellos que corresponda a «un F» o «el F». Por lo tanto, los que parecían ser términos denotativos resultan, una vez analizados, funcionar lógicamente como cuantificaciones, y éstas no son, por descontado, términos denotativos. Si digo «Hay un hombre calvo en la habitación», no me estoy refiriendo a un hombre calvo en particular, aunque lo que digo será verdadero sólo si alguien (cualquier persona) en la habitación satisface dicha descripción. Las oraciones cuantificadas son satisfechas, o no, por objetos. Si Guillermo fuera el único hombre calvo en el mundo, la calvicie de Guillermo haría verdadera la oración «alguien es calvo». Pero si en vez de ello Federico fuera el único hombre calvo en el mundo, entonces la calvicie de Federico haría verdadera esa misma oración. En cambio, una oración que contenga un término singular genuino es hecha verdadera o falsa únicamente por los estados y efectos del objeto denotado. Da igual cómo sean las demás cosas.

Aparte de su teoría de las descripciones, Russell tenía otra tesis aun más controvertida: los nombres propios ordinarios son descripciones encubiertas y, por lo tanto, cuantificaciones disimuladas. A pesar de las apariencias, los nombres propios ordinarios no son términos singulares genuinos o términos denotativos. Los únicos términos singulares genuinos son los que Russell considera nombres propios en sentido lógico, y que denotan elemen-

tos mentales corrientes conocidos infaliblemente. Por lo tanto, la categoría gramatical de términos-sujeto se subdivide en los (escasos) términos denotativos genuinos y en el resto, que en definitiva es convertible a cuantificaciones.

Éste no es el sitio apropiado para hacer un estudio de la filosofía del lenguaje de Russell de 1905. Su teoría de las descripciones es todavía aceptada por muchos, aunque no está libre de dificultades. La idea de que los nombres propios ordinarios son descripciones encubiertas es generalmente rechazada hoy (en gran parte debido a los trabajos de Saul Kripke)[12]. La razón de discutir aquí las teorías de Russell es doble. En primer lugar, éstas muestran cómo oponernos a la suposición de que cada término-sujeto con sentido se corresponde con un objeto, y nos protege así de los excesos ontológicos de Meinong. En segundo lugar, Russell, al alegar que los nombres y descripciones ordinarios son cuantificaciones encubiertas, distingue entre la gramática exterior y la estructura lógica. Por lo tanto, deja abierta la posibilidad de entender «existe» —tal como esta palabra aparece, por ejemplo, en «George Bush existe», «Superman no existe» y «los tigres existen»— como un cuantificador existencial encubierto, a pesar de que es, desde el punto de vista gramatical, un predicado.

### Bertrand Russell (1872-1970)

Bertrand Russell nació en Monmouthshire, Inglaterra, en una familia aristocrática (más tarde heredaría los títulos de conde y vizconde). Sus padres murieron siendo él un niño y fue educado por abuelos muy religiosos. Estudió en el Trinity College, donde, bajo la influencia de McTaggart, se hizo (durante un breve periodo de tiempo) idealista. Más tarde enseñó en el Trinity al mismo tiempo que G. E. Moore y tuvo a Wittgenstein como alumno. Russell escribió mucho en prácticamente todas las áreas de la filosofía, y especialmente en lógica, filosofía de la matemática, filosofía del lenguaje y de la mente y epistemología, pero también en filosofía social y política. Las obras tardías estuvieron dirigidas a un público amplio. Fue un gran estilista de la lengua inglesa, y se le concedió el Premio Nobel de Literatura en 1950. Dos de sus libros —*Los problemas de la filosofía* (1912) e *Historia de la filosofía occidental* (1945)— fueron éxitos de ventas en filosofía. Entre sus contribuciones filosóficas más importantes e influyentes se encuentran la teoría de las descripciones y la paradoja de Russell (que mostraba que el intuitivo «axioma de comprehensión» —a cada propiedad le corresponde un conjunto que contiene exactamente los elementos que poseen esa propiedad— conduce a una paradoja).

## ¿Qué es la existencia?

Al preguntar aquí «¿qué es la existencia?», nos proponemos desvelar la naturaleza de la existencia, pero el método que emplearemos será el lógico-lingüístico. Por lo tanto, considérese una oración afirmativa verdadera, tal como

«George Bush existe». Se trata, gramaticalmente hablando, de una oración dotada de un sujeto y un predicado. Una concepción —la atributiva— la toma en su valor nominal: la gramática de superficie sería la gramática real. La forma lógica de «George Bush existe» sería la aparente: Fa. «George Bush» se refiere a George Bush, el presidente estadounidense número cuarenta y uno, y «existe» se refiere al atributo o la propiedad de la existencia. George Bush tiene, así, la propiedad de la existencia. La existencia sería, pues, una propiedad de los objetos ordinarios, al igual que propiedades como el peso, la altura, la calvicie, etc.

Ahora bien, según otra concepción —la de los cuantificadores—, la gramática de superficie sería engañosa: «existe» no sería realmente un predicado sino un cuantificador. La forma lógica de «George Bush existe» no es Fa, sino $\exists x$ ($x$ = George Bush) (es decir, hay un $x$ tal que $x$ es idéntico a George Bush). Por lo tanto, la existencia no sería una propiedad de George Bush ni de ningún otro objeto ordinario[13].

Meinong y McGinn, quienes creen en objetos inexistentes, también defienden la concepción atributiva. Y no es casual que lo hagan. Si un atributo es genuino, es natural suponer que algunos objetos lo tienen y otros carecen de él. Y si la existencia es un atributo genuino, se sigue que algunos objetos existen y otros no lo hacen. Estos últimos serían los objetos inexistentes. En consecuencia, la concepción de que la existencia es un atributo, así como la concepción del significado y de la denotación, criticadas en la sección anterior, encajan perfectamente con la doctrina de que hay objetos inexistentes.

¿Cómo decidir entre la concepción atributiva y la cuantificacional? A. J. Ayer escribió una vez que si la concepción atributiva [*the property view*] fuera cierta, «se seguiría que todas las proposiciones existenciales afirmativas serían tautológicas, y todas las proposiciones existenciales negativas, autocontradictorias»[14]. No es del todo claro que el primer reparo de Ayer dé en el blanco. El hecho de señalar que la existencia es un atributo no parece implicar por sí solo que, por ejemplo, «George Bush existe» sea una tautología. Es obvio que si uno supone que todo nombre con sentido se refiere a un objeto, y que no hay objetos inexistentes, entonces el sentido del nombre «George Bush» garantizará que la oración «George Bush existe» jamás puede dejar de ser cierta. Pero en eso no interviene para nada la doctrina de que la existencia es una propiedad.

La segunda objeción de Ayer sobre el hecho de que para la concepción atributiva las proposiciones existenciales negativas son contradictorias parece tener más fuerza. Considérese la oración verdadera «Superman no existe». Según la concepción atributiva, «Superman existe» tendría la forma lógica «Fa» («a es F»); por lo tanto, «Superman no existe» tendría esa misma forma, sólo que habría que colocar «~F» (no-F) en lugar de «F»: «(~F)a» (que significa: «a es ~F»). Cualquier oración de esa forma está sujeta a la regla de la generalización existencial: de (~F)a se sigue que $\exists x$ (~F) (es decir, existe un $x$ que es no-F). Así pues, según la concepción atributiva, de la verdad de «Su-

perman no existe» se sigue que existe algo que no existe. Y esto, según Ayer dice acertadamente, sería contradictorio.

Por supuesto, hay una forma en que esta contradicción puede ser evitada. Si hay objetos inexistentes, entonces la regla de la generalización existencial no se cumple. De acuerdo con los amigos de lo inexistente, de la verdad de «Fa» sólo podemos inferir «algo es F»; y «algo es F» no implica «existe un objeto que es F». Por lo tanto, de «Superman no existe» estamos facultados para inferir sólo «hay algo que no existe». Para los defensores de lo inexistente, esto no sería una contradicción: ¡sería una verdad! Muchas cosas carecen de la propiedad de la existencia, y Superman sería una de ellas.

Sin embargo, la tesis de que hay objetos inexistentes es difícil de aceptar. En primer lugar, parece irrisorio creer que hay un ámbito de objetos que no existen. En segundo lugar, los defensores de lo inexistente deben estar dispuestos a afirmar que las expresiones castellanas «algunos» y «hay» no implican la existencia. Esto es, las oraciones «algunos Fs son G» y «hay Fs que son G» tendrían que ser entendidas de forma tal que no implicasen que existen Fs que son G. Esto es inverosímil. En tercer lugar, la suposición de que todos los nombres son denotativos (a saber, de objetos existentes o inexistentes) es particularmente inverosímil en el caso de nombres como «Vulcano». El astrónomo Jean Leverrier introdujo el nombre «Vulcano» para referirse a un planeta situado entre Mercurio y el Sol, la presencia del cual permitiría explicar ciertas observaciones astronómicas. Resultó que no había tal planeta. «Vulcano» es, pues, un caso paradigmático de ausencia denotativa, no un caso de denotación exitosa de un objeto inexistente.

Aunque la segunda objeción de Ayer en contra de la concepción atributiva quede en pie, la concepción cuantificacional tampoco sale bien parada. Según ésta, en efecto, la oración «Superman no existe» se expresa como «$\sim\exists x\ (x =$ Superman$)$» (es decir, no es el caso que haya un $x$ tal que $x$ sea idéntico a Superman). Desafortunadamente, por la regla de generalización existencial, podemos inferir: «$\exists y \sim\exists x\ (x = y)$», lo cual constituye una contradicción. Esto muestra que el denominado problema de los existenciales negativos es un problema para todos[15]. ¿Cómo podemos reconocer la verdad de «Superman no existe» sin contradecirnos a nosotros mismos o avalar lo inexistente?

¿Hay algún otro modo de decidir entre la concepción cuantificacional y la atributiva? Dos consideraciones hablan a favor de la concepción cuantificacional. Primero, dado que hemos rechazado la doctrina de los objetos inexistentes, ¿qué función desempeñaría la supuesta propiedad de la existencia? Dada la ausencia de lo inexistente, sería una propiedad de todas las cosas. Nada carecería de la propiedad de la existencia. Sin embargo, tendría que ser una propiedad puramente formal, sin efectos casuales algunos. El peso, la masa y la velocidad de un automóvil son propiedades causalmente operativas, pero la existencia del automóvil no lo sería (a pesar de que, trivialmente, el automóvil no tendría ninguna propiedad si no existiera). La postulación de una propiedad semejante parecería, pues, inútil.

En segundo lugar, incluso los defensores de la concepción atributiva reconocen que las oraciones «George Bush existe» y «$\exists x$ ($x$ = George Bush)» tienen (necesariamente) el mismo valor de verdad (aunque piensen que esta última no refleja la estructura lógica de la primera). Por consiguiente, ¿qué se ganaría con reconocer además la existencia como una de las propiedades de Bush? Ya se dijo que la existencia no cumple función explicativa alguna. ¿Por qué no suscribir el análisis cuantificacional alegando razones de economía?[16]

En suma, la concepción cuantificacional es más plausible que la atributiva. Y al rechazar la concepción atributiva deberíamos rechazar también su pariente más ilustre: la doctrina de los objetos inexistentes y la concepción de que todos los nombres y descripciones significativos denotan algo.

## Observaciones finales

En este capítulo hemos sido más condescendientes con la doctrina del realismo modal que con la de los objetos inexistentes. Algunos se asombrarán: ¿No son ambas doctrinas igual de extrañas y extravagantes? Quizás no. A favor del realismo modal de Lewis se pueden aducir varias cosas. Éste explica la verdad de nuestras aseveraciones modales y responde a interrogantes que sin él resultarían enigmáticas, como la de por qué existe nuestro mundo. En cambio, nada se puede aducir a favor de la doctrina de los objetos inexistentes, salvo el hecho de que provee de hacedores de verdad a oraciones como «Superman no existe».

Adicionalmente, hemos cuestionado dos de los motivos para creer en objetos inexistentes: la doctrina de que el significado implica la denotación y la tesis de que la existencia es una propiedad de los objetos ordinarios. Russell nos mostró la manera de rechazar la primera, y las consideraciones adelantadas en la sección precedente nos dan motivos para rechazar la segunda.

## Preguntas para el estudio

- ¿Es el realismo modal una posición sustentable?
- ¿Deberíamos asombrarnos de que exista nuestro universo?
- ¿Deberíamos admitir los objetos inexistentes?
- ¿Se te ocurre alguna objeción a la teoría russelliana de las descripciones?
- ¿Es la existencia una propiedad de Fido, así como lo son su lealtad y su temperamento jovial?

### Notas

1. Aquí prescindo de la cuestión de si existen o no objetos pasados o futuros. Véase el capítulo 5.

2. D. Lewis, *On the Plurality of Worlds* (Oxford: Basil Blackwell) 1986, pp. 2-3.
3. Esto da ocasión a críticas posibles. Lewis excluye, por definición, la posibilidad de que un único mundo contenga dos o más regiones no conectadas de forma espaciotemporal. Se podría pensar que un asunto así no debería ser zanjado de forma prescriptiva.
4. Ibíd., p. 73.
5. D. Parfit, «The Puzzle of Reality: Why Does the Universe Exist?», reimpreso en P. van Inwagen y D. Zimmerman (eds.) *Metaphysics: The Big Questions* (Oxford: Basil Blackwell) 1998, p. 419. (El artículo de Parfit va seguido de una útil réplica de R. Swinburne, pp. 427-429.)
6. Parfit, *op. cit.*, pp. 419-420.
7. S. Kripke, *Naming and Necessity* (Oxford: Basil Blackwell) 1980, p. 45, n. 13.
8. C. McGinn, *Logical Properties* (Oxford: Oxford University Press) 2000, p. 16.
9. B. Russell, «Descriptions», en su *Introduction to Mathematical Philosophy* (Nueva York: Simon & Schuster) 1961, p. 169.
10. El artículo de Russell está reimpreso en su *Logic and Knowledge* (Londres: Unwin Hyman Ltd) 1956, pp. 39-57.
11. Por lo tanto, para Russell la expresión «error denotativo» es un oxímoron. Si un témino singular gramatical puede ser significativo estando ausente su objeto, entonces no es un término denotativo. Por esta razón, Russell concluye que las descripciones definidas y los nombres propios ordinarios no son términos denotativos.
12. Véase Kripke, *op. cit.*, especialmente las conferencias I y II.
13. Según esta concepción, la existencia es una propiedad, pero de los conceptos, no de los objetos. Así, el concepto de «George Bush» tendría la propiedad de estar instanciado.
14. A. J. Ayer, *Language, Truth and Logic* (Harmondsworth: Penguin) 1976, p. 58.
15. O, al menos, un problema para cualquiera que rechace la concepción de que los nombres propios como «Superman» son descripciones encubiertas. Según la teoría de los nombres como descripciones, la oración criticada por Kripke «Superman no existe» es vertida como «$\sim\exists xFx$», donde «F» es una descripción determinada asociada con «Superman». Tal transcripción ni es contradictoria ni requiere de lo inexistente para ser verdadera.
16. Además, la concepción cuantificacional parece mejor equipada para manejar oraciones existenciales generales, tales como «los tigres existen». Esta oración es vertida como «$\exists x$ ($x$ es un tigre)», y las oraciones como «los tigres existen» expresan un mismo enunciado independientemente de cuántos tigres existan. En cambio, ¿cómo entiende la concepción atributiva «los tigres existen»? ¿De qué exactamente se predica la propiedad de la existencia? Presumiblemente, o bien de la totalidad de los tigres, o bien de cada tigre tomado individualmente. En cualquiera de estos dos casos, la oración «los tigres existen» parece tener que expresar un enunciado distinto si disminuyera el número de tigres existentes.

## Bibliografía adicional comentada

S. Kripke, *Naming and Necessity* (Oxford: Basil Blackwell) 1980. Uno de los trabajos más brillantes e influyentes de la filosofía contemporánea. Kripke propuso una filosofía del lenguaje afín al esencialismo y, con ello, rehabilitó la investigación metafísica como un área central de la filosofía.

D. Lewis, *On the Plurality of Worlds* (Oxford: Basil Blackwell) 1986. Exposición clásica y defensa del realismo modal.

C. McGinn, *Logical Properties* (Oxford: Oxford University Press) 2000. El capítulo de McGinn titulado «Existencia» brinda una exposición clara de la posición contemporánea que sigue la línea de Meinong.

## Recursos en Internet

P. Mackie (1998) «Existence», en E. Craig (ed.), *Routledge Encyclopedia of Philosophy.* Bajado el 31 de mayo de 2006 de <http://www.rep.routledge.com/article/X013>.
B. Miller (2002) «Existence», *The Stanford Encyclopedia of Philosophy (edición de verano de 2002),* Edward N. Zalta (ed.). Bajado el 31 de mayo de 2006 de <http://plato.stanford.edu/archives/sum2002/entries/existence>.
L. N. Oaklander (2005) «Negative Facts», en E. Craig (ed.), *Routledge Encyclopedia of Philosophy.* Bajado el 31 de mayo de 2006 de <http://www.rep.routledge.com/article/N118>.

# 3. Los universales y los particulares

## Introducción

El problema de los universales es uno de los más antiguos de la metafísica, y se remonta a Platón y Aristóteles. En el fondo, se refiere a la naturaleza de las propiedades y de las relaciones. ¿Son universales las propiedades y las relaciones, y, por lo tanto, son éstas idénticas en cada una de sus instancias (como pensaban Platón y Aristóteles)? ¿O podemos explicar, sin recurrir a los universales, qué significa, por ejemplo, el ser rojo para una esfera, y, en consecuencia, qué significa para dos esferas tener el mismo color?

Los filósofos que afirman que debemos apelar a los universales para explicar la naturaleza de las propiedades y de las relaciones han recibido tradicionalmente el nombre de «realistas de las propiedades»; aquellos que lo niegan, el de «nominalistas». Pero hay variedades en ambas posiciones: el realismo de las propiedades tiene versiones como la platónica, la aristotélica o la russelliana, mientras que el nominalismo se divide en nominalismo de las propiedades, nominalismo de clases, nominalismo de la semejanza y teoría de los tropos.

No es posible discutir el problema de los universales sin también discutir la naturaleza de lo que posee propiedades y es detentador de relaciones, a saber, la naturaleza de los particulares, los objetos o las **sustancias** individuales (según las distintas denominaciones otorgadas). Quienes adoptan una postura realista tradicional de las propiedades acerca de los universales respaldan la ontología del tipo sustancia-atributo (particular-universal) y, por consiguien-

te, están comprometidos con estas dos categorías del ser. Sin embargo, la noción de una sustancia individual es puesta en cuestión por los realistas de las propiedades modernos (tales como el teórico de los tropos, D. C. Williams).

## Los universales

Desde Platón, muchos filósofos han sostenido que las propiedades y las relaciones son universales, «completamente presentes» en todas sus instancias. Si mi bola de billar es roja, ello se debería, según ellos, a que la rojez (el universal) «inhiere» en la bola (el particular). Similarmente, si alguna otra bola es roja, ello se debería a que ese mismo universal «inhiere» también en esa bola. Y si una bola roja está a treinta centímetros de una bola blanca, ello se debería a que *estar a treinta centímetros de* (el universal) inhiere en la distancia espacial entre ellas. Algunas relaciones (las denominadas «relaciones internas», tales como *ser mayor que o ser más rojo que*) carecen de verdadera realidad. El que Federico sea mayor que Guillermo es algo exigido por el hecho de que Federico mide 1,90 m y Guillermo 1,70 m de alto. Mientras registremos en el inventario de la realidad las respectivas estaturas de Federico y Guillermo, no necesitamos añadir que Federico es más alto que Guillermo. Por contraste, las relaciones externas, tales como *estar situado a la izquierda de* o *hallarse a treinta centímetros de distancia de*, no vienen exigidas por sus términos, y son verdaderas adiciones a la realidad. Según algunas concepciones, las relaciones internas no son universales (ya que no son fundamentales).

## El realismo de las propiedades

Deberíamos tomar nota antes que nada de dos concepciones bastante diferentes de los universales: la platónica y la aristotélica. Según la concepción platónica, los universales son trascendentes, existen fuera del tiempo y del espacio. Son objetos **abstractos** inmutables. Pero según la concepción aristotélica, los universales son inmanentes. No existen fuera del espacio y del tiempo. Están localizados dondequiera que estén sus instancias, y sólo allí. Puesto que la existencia de entidades abstractas, fuera del espacio y el tiempo, no depende de la existencia de entidades concretas situadas en el espacio y en el tiempo, un corolario de la concepción trascendente es que los universales pueden existir aunque no estén instanciados (es decir, aunque no tengan instancias en periodo alguno de tiempo). Así, el universal «unicornio» existe, según la concepción platónica, a pesar de que no haya unicornios. Por contraste, en la concepción aristotélica, los universales no pueden existir a menos que estén instanciados.

Platón estaba interesado en las matemáticas y la geometría. Las formas geométricas ideales eran el modelo de sus universales (o «formas», como las llamó). Puesto que, por ejemplo, el teorema de Pitágoras es verdadero inde-

pendientemente de si algún ítem físico es exactamente triangular o no, los universales existen independientemente de si tienen instancias particulares o no. En cambio, Aristóteles derivó su modelo de los universales a partir del de las especies biológicas. Así como no tiene sentido pensar que pueda existir una especie de la que no hay elemento alguno, tampoco parece tener sentido que exista un universal desprovisto de alguna instancia.

### Platón (427-347 a.C.)

El fundador de la filosofía tal como la entendemos hoy, Platón, nació en el seno de una familia aristocrática de Atenas, y se convirtió en discípulo de Sócrates. Después de la muerte de éste en el año 399 a.C., Platón comenzó a escribir diálogos socráticos en memoria de su maestro. Años más tarde, fundó la Academia, la primera institución de Occidente parecida a una universidad, que proporcionó un santuario a destacados matemáticos, científicos y filósofos. Las dos doctrinas centrales de Platón fueron su teoría de las formas y su teoría sobre la inmortalidad del alma. En contraste con el mundo efímero y cambiante que vemos a nuestro alrededor, las formas serían abstractas e inmutables, auténticos objetos de conocimiento. Así, la forma de justicia existiría en un ámbito abstracto, fuera del espacio y del tiempo. Las acciones humanas individuales serían justas en virtud de participar de la forma de justicia. Todos habríamos tenido conocimiento de las formas antes de asumir nuestra actual encarnación. Las Formas son la base de la metafísica de Platón, pero desempeñan también un papel fundamental en su filosofía política. En la ciudad ideal de Platón (esbozada en su obra más famosa, *La República*), son los filósofos los que gobiernan, ya que sólo ellos entienden las Formas y, en particular, la forma del bien.

### Aristóteles (384-322 a.C.)

Nacido en la ciudad macedonia de Estagira, Aristóteles pasó gran parte de su vida adulta en Atenas. Ingresó en la Academia de Platón en el año 367 a.C., y permaneció allí dos décadas, primero como estudiante, y luego como maestro y escritor. Abandonó Atenas tras la muerte de Platón (en el 347 a.C.), pero regresó más tarde para fundar su propia escuela, el Liceo. Aristóteles fue el filósofo más sistemático de la antigüedad. Fue el primero en concebir la filosofía como una investigación intelectual dividida en distintas ramas del saber: la lógica, la investigación científica (integrada sobre todo por la biología y la cosmología), la metafísica, el problema mente-cuerpo, la ética y la política, así como la crítica literaria. Aristóteles hizo importantes contribuciones en todos estos campos y, al hacerlo, ayudó a definir el ámbito de lo que hoy entendemos como filosofía. Fue muy estudiado por los pensadores medievales, sobre todo tras la aparición en el siglo XII de una serie de traducciones de sus obras. Su influencia se incrementó y legitimó gracias a que santo Tomás de Aquino había convertido al aristotelismo en la base de su teología católica.

La esencia del realismo tradicional de las propiedades es razonablemente clara: los objetos (los particulares o «sustancias individuales», según el vocabulario más tradicional) tienen, es decir, instancializan, propiedades; dos (o más) objetos pueden, literalmente hablando, tener una y la misma propiedad; por lo tanto, las propiedades no son sino universales que pueden estar completamente presentes en dos o más lugares al mismo tiempo[1].

Un motivo para suscribir el realismo de las propiedades —motivo que, ciertamente, fue uno de los que inspiraron a Platón— procede de consideraciones relativas al significado de los términos generales. Supóngase que empezamos asumiendo una teoría denotativa del significado, de acuerdo con la cual el significado de una palabra es una entidad, a saber, la entidad denotada por la palabra. En el caso de términos singulares como, por ejemplo, los nombres propios ordinarios «Sócrates», «Red Rum» y «Edimburgo», no es difícil identificar los entes que, de acuerdo con la teoría denotativa, serían el significado de las palabras correspondientes: el hombre Sócrates, el caballo Red Rum y la ciudad de Edimburgo, respectivamente. Pero ¿qué pasa con términos generales como «caballo» y «ciudad», es decir, con términos que se aplican a muchos entes? ¿A qué entes se referirían esas palabras? No, por cierto, a un caballo particular o a una ciudad particular (pues, ¿por qué un caballo o una ciudad determinados?). De ahí que se los denomine términos generales: se aplican de manera completamente general a todos los caballos y a todas las ciudades, respectivamente. Por lo tanto, pensó Platón, si los términos generales no designan particulares, deben designar universales (como, digamos, la *urbanidad* o la *equinidad*).

Hoy pocos filósofos aceptan lo anterior como un argumento a favor de los universales. En primer lugar, el argumento presupone que si «caballo» denota algo, denota o bien un caballo particular o bien el universal *equinidad*. Pero esta suposición es cuestionable: ¿por qué no suponer que «caballo» se refiere a cada caballo particular? Un defensor del argumento semántico debería descartar esto último si quiere que su argumento tenga éxito. En segundo lugar, y de manera más fundamental, ¿Por qué deberíamos aceptar la idea subyacente de que el significado de una palabra es cierta entidad representada por esa palabra? (Véase el capítulo 2.) Nombres como «Papá Noel» y «Odiseo», por ejemplo, están totalmente dotados de sentido y, sin embargo, no denotan a persona alguna (son designaciones vacías). En lugar de concebir el significado como denotación, deberíamos quizás pensar en el significado de una palabra como una función de sus usos en el seno de una comunidad lingüística. Si así lo hiciéramos, el referente de una palabra, si es que tiene alguno, sería irrelevante para su significado.

No obstante, hay otro argumento, esta vez metafísico, a favor de los universales. Reza como sigue:

Considérense dos esferas exactamente similares. Tienen el mismo color (entre otras similitudes). Esto es, el color de una es literalmente el mismo que el color de la otra, es decir, es **numéricamente idéntico** a él. Lo que está presente en una de ellas también está presente

en la otra. Los particulares no pueden estar (completamente) en dos lugares al mismo tiempo, pero los universales sí. Ningún particular (como un hombre o un caballo) puede estar enteramente presente en dos lugares al mismo tiempo. Por supuesto, una parte de un caballo (por ejemplo, su pata delantera izquierda) puede y tendrá que estar en un lugar diferente que otra de sus partes (por ejemplo, su pata delantera derecha). Pero el caballo no puede, todo él, estar en dos lugares al mismo tiempo. Por contraste, la esencia de un universal es su repetibilidad: puede estar completamente presente en lugares diferentes al mismo tiempo. Así, para explicar el truismo de que diferentes objetos pueden tener la misma propiedad, enteramente presente en cada objeto, debemos apelar a los universales.

El problema con este argumento es que «lo mismo» no se usa siempre para expresar una estricta identidad numérica. A veces sí, como cuando, por ejemplo, decimos que dos niños «tienen la misma madre». En este caso, es indudable que queremos decir que la madre de uno de ellos es, literal y numéricamente, la misma que la madre del otro. Pero considérese el uso de «lo mismo» en «él tiene los mismos ojos que su padre». Aquí la palabra «mismos» expresa una **identidad cualitativa** (esto es, una similitud sorprendente), no una identidad numérica. El sentido de «lo mismo» en «tienen la misma madre» es totalmente diferente del sentido que la expresión adopta en «tienen los mismos ojos».

Esta observación rebate el argumento metafísico empleado a favor de los universales. Del hecho de que X es el mismo F que Y no se sigue que hay alguna propiedad de X que es numéricamente idéntica (a diferencia de meramente similar) a una propiedad de Y. El defensor del argumento metafísico tendría que demostrar que «el mismo F» es usado aquí para expresar una identidad numérica. Puede que lo consiga en relación con términos generales sustantivos (como «hombre» o «caballo»), pero será más difícil que lo haga en relación con términos generales cualitativos (como «rojo» o «redondo»).

## Objeciones al realismo tradicional de las propiedades

Ni el argumento semántico ni el metafísico a favor de los universales son, pues, convincentes. ¿Hay, aparte de ellos, algún argumento positivo en contra de los universales? Una objeción al realismo tradicional de las propiedades es que no ofrece explicación alguna de la conexión existente entre un particular (por ejemplo, Sócrates) y sus propiedades. Se dice que un particular instancializa a varios universales. Pero ¿qué es la instanciación? ¿Es acaso una relación? ¿Es primitiva e inanalizable? ¿No debería parecernos misteriosa?[2]

En segundo lugar, la explicación de la comunidad de algo [*commonality*] en términos del mismo universal conduce a un regreso al infinito. Aristóteles adelantó esta objeción contra la concepción platónica y trascendente de los universales. Su «argumento del tercer hombre» se proponía demostrar que la teoría platónica de las formas implica un regreso al infinito. (Este argumento

se puede hallar ya en el *Parménides* de Platón, pero la formulación de Aristóteles es más conocida.)

Partiendo del supuesto de que la forma F es ella misma un F (el denominado supuesto de la autopredicabilidad), el argumento del tercer hombre procede de la siguiente manera: Platón intenta explicar lo que todos los hombres individuales tienen en común, aquello en virtud de lo cual son subsumibles bajo el término general «hombre», apelando a una relación («participación» o «imitación») en que cada hombre se encuentra con respecto a la forma hombre. Pero si la forma hombre cae ella misma bajo el término general «hombre», entonces tendremos necesidad de postular una forma adicional $hombre_1$ para poder explicar lo que todos los hombres individuales y la forma hombre tienen en común. Y si la forma $hombre_1$ es ella misma un hombre, entonces tendremos que postular una tercera forma, $hombre_2$, para explicar por qué los hombres individuales, la forma hombre y la forma $hombre_1$ son todos hombres. Y así, *ad infinitum*. El regreso es vicioso porque para explicar el compartir algo a cualquier nivel estamos siempre forzados a remontarnos al nivel siguiente, de modo que jamás alcanzamos explicación genuina alguna.

Aunque el argumento aristotélico del tercer hombre estaba dirigido contra la teoría platónica de las formas, muchos han pensado que proporciona un argumento contra todas las teorías de los universales. Incluso aquellas teorías de los universales como la aristotélica, que no parten del supuesto de la autopredicabilidad y conciben los universales como inmanentes al mundo, serían vulnerables al siguiente argumento de regreso al infinito: cuando dos o más particulares instancializan un universal, esas instanciaciones serán ellas mismas instancias del universal instanciación. Lo que explica que distintas instancias de instanciación compartan algo deberá ser la presencia de un universal ulterior $instanciación_1$. Y lo que explica lo que hay de común entre las distintas instancias de $instanciación_1$ deberá ser la presencia de un universal ulterior $instanciación_2$. Y así, *ad infinitum,* si admitimos el supuesto del realista de que cualquier comunidad de algo debe ser explicada como la instanciación de un universal.

Podemos presentar el regreso en términos menos tradicionales del modo siguiente. Si A y B tienen algo en común (digamos, si ambos son F), el realista postula un universal (la F.dad) que ambos instancializan. Pero entonces el hecho o estado de cosas de que A instancializa la F.dad, y el de que B instancializa la F.dad, tienen algo en común: ambos son instanciaciones; así es que debemos postular un universal (instanciación) que ambos estados de cosas instancialicen[3]. Pero ahora los dos estados de cosas complejos, a saber, que {[A instancializa la F.dad] sea una instancia de la instanciación}, y que {[B instancializa la F.dad] sea una instancia de instanciación} tendrían algo en común; así es que debemos postular un universal ulterior ($instanciación_1$) que ambos estados de cosas instancialicen. Y así sucesivamente.

En tercer lugar, hay un problema que concierne a lo que posee propiedades o instancializa universales, esto es, un problema que concierne al particu-

lar o a la sustancia individual. El realista de las propiedades tradicional endosa la ontología convencional de sustancia y atributos (que corresponde a la distinción gramatical entre sujeto y predicado). Una sustancia (como un caballo particular) no se identifica con sus propiedades, sino que las *tiene*. Una sustancia no es sólo la suma de sus propiedades; es algo que se suma a, o que se resta de, sus propiedades. Pero ello suscita una inmediata duda epistémica: puesto que todo lo que percibimos y provoca nuestra reacción son las propiedades, ¿cómo podemos acceder a las sustancias?

## La teoría russelliana del racimo

Bertrand Russell (1872-1970) fue un realista de las propiedades que rechazaba la noción tradicional de sustancia y consideraba los objetos ordinarios como racimos [*bundles*] de universales. Trataba de esquivar de ese modo algunas objeciones convencionales al realismo tradicional de las propiedades, como las que acabamos de formular. Por ejemplo, no habrá nada enigmático en el vínculo entre la sustancia y el universal, si no hay sustancias (tal como se las concibe normalmente).

¿Qué le disgustaba a Russell en la noción de sustancia? Una sustancia o particular, escribió:

> no puede ser definida, identificada o conocida; es únicamente algo al servicio de la función meramente gramatical de proporcionar un sujeto en una oración sujeto-predicativa como «esto es rojo». Y se considera hoy que permitir que la gramática dicte nuestra metafísica es una empresa arriesgada. [...] La noción de una sustancia como una percha de la que colgar predicados es repugnante[4].

La respuesta de Russell consistió en eliminar la categoría de sustancia y concebir los objetos ordinarios como racimos de universales. Una vez que hubo prescindido de las sustancias, necesitaba dar cuenta de qué es lo que unifica o separa distintos objetos. Para explicar la unidad de un objeto singular, recurrió a la noción de «copresencia» [*compresence*], definida inicialmente para el caso de experiencias. Si observo cómo un pájaro pasa volando y simultáneamente oigo el repicar de una campana, estas dos experiencias estarán copresentes (serán percibidas por el mismo sujeto). Más generalmente, las cualidades que, estando unidas entre sí, constituyen un objeto, como mi silla, estarán copresentes unas con otras, de manera tal que ninguna otra cualidad en el universo esté copresente con todas esas cualidades. Esas cualidades son, para Russell, los universales: la cualidad de lo marrón de mi alfombra es estrictamente idéntica a la cualidad de lo marrón de mi escritorio.

Sin embargo, la teoría de Russell adolece de los dos problemas siguientes. En primer lugar, la «copresencia» es sólo una etiqueta. No se nos dice cómo varios universales se las arreglan para formar juntos, sin ser inherentes a una

sustancia, un objeto, tal como un caballo particular. En segundo lugar, ¿qué es lo que, según Russell, hace que los objetos sean distintos entre sí? El propio Russell reconoce que, aunque los hombres tienen algunas cualidades en común (tales como la humanidad), no hay dos hombres exactamente iguales entre sí. Siempre se diferenciarán en algo, por muy trivial que sea (por ejemplo, diferencias en el número de sus cabellos, o en el número de moléculas que componen sendas manos izquierdas, etc.). Como dice Russell, «lo que hace que una instancia sea única es sólo la combinación de cualidades»[5]. Así, Russell puede aseverar que cada vez que dos individuos son distintos habrá diferencias en los correspondientes racimos de universales en virtud de los cuales son distintos.

Pero esta afirmación sólo consigue despertar la siguiente objeción: aunque no haya dos objetos exactamente iguales, ¿no sería posible que los hubiera? Por ejemplo, ¿no podría existir un mundo que contuviera dos esferas exactamente similares? Sin embargo, para Russell, tal mundo no sería posible. Dado que los objetos son racimos de universales, la identidad de dos racimos implica la identidad de los objetos. Si las «dos» esferas están compuestas de los mismos universales, no pueden ser dos en número. «Ellas» son una. Pero, intuitivamente, sería posible que hubiera dos esferas exactamente similares. La teoría de Russell excluye esta posibilidad, y eso es una razón para criticarla.

## Variedades del nominalismo

Hemos visto que hay complicaciones con el realismo tradicional de las propiedades y la teoría russelliana del racimo. ¿Son las perspectivas de éxito del nominalismo más halagüeñas? Examinemos la gama de opiniones que intentan explicar los objetos y sus propiedades sin recurrir a los universales.

### El nominalismo de los predicados

De acuerdo con el nominalismo de las predicados, el que $x$ sea F equivale a que el predicado «F» se aplique a $x$, o sea verdadero de $x$. Y lo que hace verdadero que $x$ e $y$ sean ambas F es que «F» se aplique tanto a $x$ como a $y$. Así, lo que hace que sea verdadero que una esfera sea roja es simplemente el que el predicado «rojo» se aplique a esa esfera.

Esta teoría es vulnerable a objeciones severas. En primer lugar, en el universo hay con toda seguridad propiedades que nunca descubriremos, y para las cuales no existe un predicado del lenguaje natural. El nominalismo de los predicados parece esencialmente incapaz de admitir la posibilidad de tales propiedades. En segundo lugar, y más fundamentalmente, el nominalismo de los predicados parece poner el carro delante de los bueyes. Pues es obvio que

el que *x* sea F no consiste en aplicar «F» a *x;* al contrario, «F» se aplica a *x* porque *x* es F. Si *x* es F (digamos, una montaña de una altura determinada), *x* sería F incluso si no existiera lenguaje alguno. (Estas objeciones también se aplican al nominalismo de los conceptos, o sea, a la idea de que *x* es F en virtud del hecho de que *x* cae bajo el concepto de F.)

## El nominalismo de las clases

Para el nominalismo de las clases, el que *x* sea F equivale a que *x* pertenezca a la clase de los F. Y lo que hace verdadero que *x* e *y* sean ambas F es que tanto *x* como *y* sean miembros de la clase de los F. Así, el ser rojo es para una esfera el que ésta pertenezca a la clase de las cosas rojas.

El nominalismo de las clases evita la primera objeción al nominalismo de las clases. El que un objeto sea miembro o no de una clase determinada no depende de si nosotros hemos descubierto dicha clase o poseemos un nombre para designarla. (La clase de los electrones existió antes de que la conociéramos.) Sin embargo, la segunda objeción parece seguir en pie: es obvio que *x* es miembro de la clase de los F porque *x* es F, y no al revés.

Además, el nominalismo de las clases adolece de dos dificultades propias. En primer lugar, la relación de pertenecer a una clase es ella misma un universal, que se instancializa cada vez que un objeto es miembro de una clase. Así pues, el nominalismo de las clases depende implícitamente de los universales, y, por ende, no consigue ser una versión genuina del nominalismo. En segundo lugar, supóngase que todos los F, y nada más que los F, son G (por ejemplo, supóngase que todas las cosas rojas, y sólo las cosas rojas, son redondas). Dado que la clase de los F es la clase de los G, se sigue, según el nominalismo de las clases, que la propiedad de ser F es idéntica a la propiedad de ser G: un resultado ciertamente absurdo.

Un problema relacionado con el anterior es el que surge para cualquier par de términos generales vacíos, tales como «unicornio» y «dragón». Al aplicar la estrategia nominalista, obtenemos lo siguiente: *x* es un unicornio si *x* pertenece a la clase de los unicornios, e *y* es un dragón si *y* pertenece a la clase de los dragones. Sin embargo, la clase de los dragones es la misma que la clase de los unicornios (esto es, la clase sin elementos, la clase vacía). En cuyo caso, el nominalista de las clases se ve llevado a la conclusión absurda de que la propiedad de ser un unicornio es la misma que la propiedad de ser un dragón. (Esto es una variante de la segunda objeción, ya que si no hay F y no hay G, entonces se sigue trivialmente que todos los F, y sólo los F, son G.)

## El nominalismo de la semejanza

El nominalismo de la semejanza estipula que el que *x* sea F equivale a que *x* sea elemento de una clase de objetos que se asemeja a otra, en que «semejanza» es tratada como una relación primitiva e inanalizable. Lo que hace verdadero que *x* e *y* sean ambas F es que se asemejan entre sí. Una esfera es roja porque es un elemento de una clase de objetos semejantes entre sí.

Sin embargo, hay tres problemas con esta versión del nominalismo. En primer lugar, un objeto como una esfera roja es elemento de toda una serie de clases de semejanza. Por ejemplo, es elemento de la clase de las cosas rojas, y elemento de la clase de las cosas esféricas. Ahora bien, es claro que la esfera no es roja por el hecho de que se asemeje a las cosas esféricas; es roja en virtud de que se asemeja a las cosas rojas. Esto es, es roja porque se asemeja a otros objetos con respecto al color (y no con respecto a la forma). Pero entonces, la semejanza deja de ser una relación inanalizable, y admite modalidades. Y ¿qué son las modalidades sino universales?

Russell tuvo una objeción parecida, pero más fundamental: la relación de semejanza es ella misma un universal. Escribió:

> Si deseamos evitar los universales *blancura* y *triangularidad,* escogeremos algún parche particular de blanco y algún triángulo particular, y diremos que cualquier cosa es blanca o triangular si tiene el tipo correcto de semejanza con el particular respectivamente seleccionado. Pero entonces la semejanza requerida tendrá que ser un universal. Debido a que hay muchas cosas blancas, la semejanza tendrá que darse entre muchos pares de cosas blancas particulares; y eso era precisamente lo que caracterizaba a un universal[6].

En este caso, el nominalismo de semejanza, al igual que el nominalismo de clases, no habría podido evitar un compromiso implícito con los universales.

Finalmente, hay un problema con los términos generales que poseen una única instancia. De acuerdo con el nominalismo de semejanza, un objeto es esférico si se asemeja a otros objetos en relación con la forma. Pero ¿qué pasaría si el universo hubiera contenido una sola esfera y nada más? Seguramente ésta habría sido esférica. Y, sin embargo, el nominalista de la semejanza tendría que negarlo, ya que no habría en el mundo algo a lo que la esfera pudiera asemejarse. Esto es, ciertamente, una consecuencia inadmisible.

## Los tropos

El nominalismo de los tropos es la versión más interesante y radical del nominalismo. Todas las otras versiones del nominalismo están de acuerdo con el realismo de las propiedades acerca de qué cosas son particulares; sólo discre-

pan en relación con la eventual necesidad de postular entidades ulteriores, los universales. La teoría de los tropos difiere tanto del nominalismo moderado como del realismo de las propiedades en cuanto a la identificación de los particulares: concibe, esto es, las propiedades y las relaciones como particulares. Los tropos (es decir, las propiedades y las relaciones concebidas como particulares) han sido denominados «particulares abstractos», en contraste con particulares más conocidos (los objetos concretos ordinarios)[7]. Así, tú, yo y la Torre Eiffel son particulares concretos; tu sonrisa, el color de mis ojos y la altura de la Torre Eiffel son particulares abstractos.

Aunque formulada en el siglo XIX por G. F. Stout, la teoría de los tropos fue considerablemente desarrollada y defendida en el siglo XX por Donald Williams y Keith Campbell. Para Williams, los tropos son los ítems básicos del universo, el «alfabeto del ser». «Cualquier mundo posible, y, por lo tanto, también el nuestro, está completamente constituido por sus tropos y las conexiones de localización y semejanza entre éstos.»[8] De acuerdo con la teoría de los tropos, las propiedades y las relaciones son particulares abstractos, no universales. Así, supóngase que tenemos dos bolas de billar a la vista. Son completamente semejantes. En especial, ambas son rojas, pero no porque compartan una propiedad común. Más bien, la rojez de una de las bolas es un particular (irrepetible), numéricamente distinta de, pero exactamente semejante a, la rojez de la otra bola. Cada rojez es un particular abstracto distinto (un tropo).

Normalmente, contrastamos un particular concreto con sus propiedades; por ejemplo, contrastamos un hombre (Guillermo) con su calvicie. No pensamos que la calvicie de Bill sea ella misma un particular; pero eso es precisamente lo que la teoría de los tropos afirma. Otros podrán ser igual de altos que Guillermo, pero tienen su propio tropo de estatura, numéricamente distinto del de Guillermo. Las relaciones también son tropos: el que Guillermo sea treinta centímetros más alto que Federico, y el que Ana sea treinta centímetros más alta que María, no son la misma relación: son dos tropos relacionales exactamente similares.

Según la teoría de tropos convencional, los objetos concretos ordinarios no son sustancias (tal como éstas son entendidas normalmente), sino racimos de tropos. ¿Qué relación habría entre un objeto concreto y los tropos que lo constituyen? Si hemos de hacer justicia a la intuición de sentido común respecto de que algunas propiedades de un particular concreto son accidentales, la relación de un objeto concreto con sus tropos no puede ser la misma que la de un conjunto con sus elementos. Dado que la calvicie es una de las propiedades **accidentales** de Guillermo, no deberíamos pensar en la calvicie de Guillermo como un elemento del conjunto de tropos que constituyen a Guillermo. Si lo hiciéramos, se seguiría, debido a que un conjunto tiene a sus elementos esencialmente, que Guillermo es esencialmente calvo (lo cual no es cierto). Por contraste, otras propiedades de Guillermo podrían ser **esenciales** a él (por ejemplo, su humanidad). Quizás sea imposible para Guillermo

existir de otra manera que siendo un ser humano. Si así fuera, entonces el tropo de humanidad de Guillermo sería esencial a ese racimo que es Guillermo.

La relación entre un objeto ordinario y sus tropos debe presumiblemente ser del tipo todo-parte, no del tipo conjunto-elemento. Algunas partes son accidentales, y otras esenciales. El misterioso «es» de la instanciación resulta, de ese modo, reemplazado por el «es» de «es una parte de». La oración «Guillermo es calvo» es verdadera no porque Guillermo instancialice el universal *calvicie*, sino porque un tropo de calvicie es parte del racimo de tropos que es Guillermo.

El gran mérito de la teoría de los tropos es que evita muchas de las objeciones que han sido esbozadas hasta ahora. Evita las objeciones a otras versiones del nominalismo porque no analiza lo que significa que un objeto tenga una propiedad en términos de una relación entre ese objeto y otra cosa (como un predicado, una clase, una clase de objetos semejantes, etc.). Si una esfera es roja es porque contiene un tropo rojo, y esto no tiene nada que ver con que la esfera se encuentre en una relación con alguna otra cosa. Aunque la teoría de los tropos accede por su cuenta a la noción de semejanza, no concibe esta relación, ni ninguna otra, como un universal. El que A se asemeje a B y el que C se asemeje a D son tropos de semejanza distintos, no un universal doblemente instancializado. Así, los problemas que destacamos con respecto al realismo de las propiedades no se plantean para la teoría de los tropos. Por ejemplo, no hay peligro alguno de un regreso al infinito, ya que el compartir algo no se explica en términos de instanciación, sino en términos de semejanza.

Sin embargo, la teoría de los tropos adolece de algunos problemas. En primer lugar, los teóricos de los tropos reemplazan la relación de instanciación por la relación de la parte y el todo. Pero ¿en qué consiste esa relación? No se nos ha dicho nada acerca de ella, al margen de que algunas partes (tropos) son esenciales para el racimo, y otras, en cambio, accidentales. En lugar de esclarecer las cosas, esto suscita otro interrogante: ¿Cómo puede ser un tropo esencial para un racimo? ¿Qué hace que sea esencial? ¿Es un racimo el tipo de cosa que puede tener partes esenciales?

En segundo lugar, incluso si no se le plantease al teórico de los tropos el problema de la instanciación, éste tendría que responder a la pregunta de qué es lo que vincula entre sí a los tropos para formar un objeto singular. (Russell, aunque utilizaba en su teoría una noción de racimo distinta, tenía un problema análogo.) Un racimo de tropos podrá estar «co-localizado», pero, ¿cuál sería el cemento que une a éstos? ¿Pertenece esta pregunta a la ciencia o a la metafísica? ¿Puede ser respondida acaso?

En tercer lugar, la estrategia para el teórico de los tropos es sustituir la relación de instanciación por la relación de la parte y el todo. Esto funciona bastante bien para el caso de las predicaciones que involucran a particulares concretos, pero ¿qué pasa con aquellas que involucran tropos? Podemos decir muchas cosas verdaderas acerca de un tropo de la rojez: que es rojo, que tiene color, que es rojo o es verde, que es mi tropo favorito, que se halla en Cambe-

rra, que ha durado todo el 2006, etc. Dado que un tropo no es un racimo, no podemos considerar «este tropo es rojo» como algo cuya verdad esté determinada por un racimo que contiene un tropo de rojez. Entonces, ¿cómo entender estas predicaciones, si no es en términos de la instanciación de una propiedad en un tropo? Y ¿no nos remite eso de nuevo a los universales?

Hay problemas con la teoría de los tropos, así como con otras teorías. Sin embargo, es una explicación original de la naturaleza de los objetos y las propiedades, digna de investigación ulterior.

## Los tropos

La teoría de los tropos es una versión radical del nominalismo. Como todas las versiones del nominalismo, niega la existencia de los universales. Adicionalmente, la teoría de los tropos afirma la idea radical de que las propiedades y las relaciones son ellas mismas particulares, a saber, particulares abstractos. Así, la rojez de una bola de billar concreta es un particular abstracto, localizado donde se encuentra la bola y en ningún otro lugar. Una bola de billar distinta, aunque se asemeje exactamente a la primera, tendrá un tropo de rojez numéricamente distinto, aunque sea exactamente semejante al primero. No hay una propiedad de color que sea común a, o esté instanciada en, ambas bolas. Algo similar sucede con todas las demás propiedades de las bolas y con las relaciones que se dan entre ellas. Según la versión estándar de la teoría de los tropos, los particulares ordinarios (esto es, lo que todas las teorías clasifican como particulares: tú, yo, el siguiente ganador de la Copa de Melbourne, etc.) están compuestos de particulares abstractos. Una bola de billar está compuesta de un tropo de color, un tropo de forma, un tropo de peso, etc. El universo todo consiste en tropos. Son el «alfabeto de la realidad», del cual está construido todo lo demás. Quedan abiertas, sin embargo, algunas preguntas en relación con la teoría de los tropos: ¿Podría un tropo de la rojez existir por sí solo? Si no es así, ¿por qué no? ¿Qué unifica a tropos distintos para formar objetos individuales? ¿Son básicos algunos tropos, y otros, en cambio, reducibles o eliminables?

## Observaciones finales

Hemos cubierto mucho terreno difícil en este capítulo, y nuestras conclusiones son las siguientes: tanto el realismo de las propiedades tradicional como el russelliano se enfrentan a problemas serios, y lo mismo les sucede a las explicaciones nominalistas, tanto moderadas como extremas. De las teorías disponibles, la de los tropos podría ser la explicación de las propiedades más prometedora, o quizá la menos insatisfactoria, aunque se trate de un proyecto inconcluso que aún tiene muchas aristas que limar.

## Preguntas para el estudio

- ¿Qué diferencia hay entre la concepción platónica y la concepción aristotélica de los universales?
- ¿En qué consiste la distinción entre la identidad numérica y la cualitativa?
- ¿Cómo podría un defensor de los universales neutralizar el regreso al infinito implicado por la instanciación?
- ¿Se puede defender exitosamente alguna versión del nominalismo?
- ¿Qué es un tropo?

### Notas

1. Los universales a menudo se distinguen de los particulares por su capacidad para «estar completamente presentes en dos o más lugares al mismo tiempo». Pero esto no puede ser del todo correcto si hay particulares que existen fuera del espacio y del tiempo. Según ciertas concepciones, los números se encontrarían fuera del espacio y del tiempo, instancializando universales: por ejemplo, cada número par instancializaría el universal consistente en ser múltiplo de dos. Pero este universal no está presente completamente en dos lugares al mismo tiempo, ya que los números no existen en el espacio y el tiempo. Esta consideración muestra que los conceptos de «instanciación múltiple» y «localización múltiple» no son idénticos entre sí.
2. Si ya es misterioso cómo un particular instancializa a un universal, aún lo es más la forma en que un universal de orden inferior instancializa a un universal de orden superior (como *rojo* instancializa *color*). Cabría preguntarse si la noción de instanciación es en ambos casos la misma.
3. Hablar de estados de cosas no es una simple *façon de parler.* Un realista tradicional de las propiedades concibe el estado de cosas por el que A es F como una adición a la ontología, más allá de los simples A y F: un mundo podría contener A y F sin por ello tener que incluir el hecho de que A sea F, si A no fuera F en ese mundo, aunque otras cosas lo fueran. Y, por ello mismo, el plus requerido no sería otra entidad, digamos, una que correspondiera al «es» de la instanciación. (Véase D. M. Armstrong, *Truth and Truthmakers* [Cambridge: Cambridge University Press] 2004, cap. 4.)
4. B. Russell, *Human Knowledge, its Scope and Limits* (Nueva York: Simon & Schuster) 1948. Reimpreso en *Metaphysics: The Big Questions,* ed. por P. van Inwagen y D. Zimmerman (Oxford: Basil Blackwell) 1999, p. 54.
5. Ibíd., p. 57.
6. B. Russell, *The Problems of Philosophy* (Oxford: Oxford University Press) 1978, p. 55.
7. Aquí «abstracto» no significa «fuera del espacio y del tiempo», sino «tenue», «sutil» o «difuso». Véase D. C. Williams, «On the Elements of Being: 1», en D. H. Mellor y A. Oliver (eds.), *Properties* (Oxford: Oxford University Press) 1997, p. 115. Véase también K. Campbell, *Abstract Particulars* (Oxford: Basil Blackwell) 1990.
8. Williams, *op. cit.*, p. 116.

## Bibliografía adicional comentada

D. M. Armstrong, *Universals: An Opinionated Introduction* (Boulder, Col.: Westview Press) 1989. Con mucho, la mejor introducción al problema de los universales, escrita por uno de sus más enconados defensores. Armstrong incluso apela a los universales para explicar las leyes de la naturaleza: el que sea una ley de la naturaleza que todos los F son G se debe a que los universales F y G están necesariamente conectados.

K. Campbell, *Abstract Particulars* (Oxford: Basil Blackwell) 1990. Una clara y mordaz defensa de los tropos.

B. Russell, *The Problems of Philosophy* (Oxford: Oxford University Press) 1978, capítulos 9 y 10. Escrito en 1912, este libro sigue siendo una de las introducciones clásicas a la filosofía. En el capítulo 9, Russell arguye a favor de la existencia de los universales (reconstruidos de forma trascendental), y adelanta una célebre objeción contra el nominalismo de la semejanza. En el capítulo 10 considera de qué forma conocemos los universales. Afirma que las cualidades sensibles y las relaciones pueden ser conocidas directamente, por familiaridad [*acquaintance*], y que todo conocimiento a priori es un conocimiento de relaciones entre universales.

## Recursos en Internet

J. C. Bigelow (1998) «Universals», en E. Craig (ed.), *Routledge Encyclopedia of Philosophy.* Bajado el 31 de mayo de 2006 de <http://www.rep.routledge.com/article/N065>.

C. Daly (2005) «Properties», en E. Craig (ed.), *Routledge Encyclopedia of Philosophy.* Bajado el 31 de mayo de 2006 de <http://www.rep.routledge.com/article/N121SECT4>.

C. Swoyer, «Properties», *The Stanford Encyclopedia of Philosophy (edición de invierno de 2000),* Edward N. Zalta (ed.). Bajado el 31 de mayo de 2006 de <http://plato.stanford.edu/archives/win2000/entries/properties>.

# 4. La causalidad

## Introducción

Nuestro lenguaje está repleto de frases etiológicas, es decir, frases que connotan causalidad. A menudo usamos la palabra «causa» o sus derivados —como por ejemplo, en «fumar produce cáncer», «sus palabras causaron malestar», etc.—, pero también recurrimos a muchos verbos que presuponen causalidad: «ellos tuvieron que empujar su coche», «ella lo abrazó», etc. Estas locuciones suscitan un gran número de interrogantes sobre cuya respuesta no hay consenso. La más importante de ellas se refiere a la naturaleza misma de la causalidad. La discusión moderna de esta cuestión fue inaugurada por David Hume (1711-1776). Pero también hay cuestiones preliminares sobre la causalidad que deberían ser abordadas primero, y que permiten preparar la escena para la pregunta constitutiva más familiar sobre la naturaleza de la causalidad.

## Cuestiones preliminares

He aquí seis cuestiones preliminares:

(1) ¿Hay dos tipos de causalidad (la singular y la general), o, más bien, una sola clase de causalidad, que puede tener lugar ya entre «signos-tipo» [*types*], y entre «signos-ocurrencia» [*tokens*]?

## David Hume (1711-1776)

Nacido en Berwickshire, Escocia, Hume es considerado generalmente el filósofo anglófono más importante que ha existido. También fue un notable historiador y ensayista. Heredero intelectual de Locke y Berkeley, Hume fue un empirista, un naturalista y un escéptico. Sus obras filosóficas más importantes —*Tratado de la naturaleza humana* (1739-1740), *Investigación sobre el entendimiento humano* (1748), *Investigación sobre los principios de la moral* (1751), así como los *Diálogos sobre la religión natural* (1779), publicados póstumamente— fueron tremendamente influyentes, aunque criticadas en su tiempo como escépticas y ateas. La fama y la fortuna que Hume llegó a adquirir en vida se debieron en gran parte a la publicación de su *Historia de Inglaterra* en seis volúmenes (1754-1762). Su reputación de ateo evitó que fuera nombrado para cátedras en Edimburgo (1745) y Glasgow (1752). Nunca desempeñó cargo académico alguno. Hombre amable y conversador, Hume frecuentó a destacados intelectuales de su tiempo. Pudo contar entre sus amigos y conocidos a Adam Smith, James Boswell, Denis Diderot y Jean-Jacques Rousseau.

(2) ¿Cuáles son los correlatos [*relata*] de una relación causal? Esto es, ¿qué clases de entidad son A y B cuando decimos algo de la forma «A causó B»?
(3) ¿Qué lógica tiene la relación causal?
(4) ¿Qué dirección temporal tiene la causalidad?
(5) ¿Con qué derecho calificamos a una determinada causa, de entre una plétora de circunstancias dadas, como «la» causa de un evento?
(6) ¿Qué conexión hay entre la causalidad y las leyes de la naturaleza?

(1) Vale la pena distinguir entre **aseveraciones causales singulares** («sus palabras causaron molestia») y **aseveraciones causales generales** («el fumar causa cáncer»). Las primeras conectan entre sí fenómenos particulares fechables (signos-ocurrencia [*tokens*]), mientras que las segundas ponen en relación tipos de fenómenos (signos-tipo [*types*]). Por supuesto, es prácticamente imposible que las afirmaciones particulares no estén vinculadas con las generales. El hecho de que fumar sea causante del cáncer no puede no estar vinculado con el hecho de que bocanadas de humo particulares hagan que individuos concretos sean afectados por el cáncer. No obstante, queda por dilucidar si, además de afirmaciones singulares y generales, hay tipos singulares y generales de causalidad. Cualquiera que sea la respuesta, en este capítulo nos ocuparemos únicamente de la causalidad singular.

(2) ¿Cuáles son los correlatos en casos de causalidad (singular)? Tres candidatos estándar son: los objetos [*objects*], los eventos [*events*] y los he-

chos [*facts*]. En cuanto a los primeros, decimos, efectivamente, cosas como «el coche mató al hombre», «María mató a Sara», etc., dando a entender que hay determinados objetos (el coche y el hombre, María y Sara) que son los correlatos de la relación causal «mató». Ciertamente, se podría argumentar que esto es sólo una manera imprecisa de hablar. Pues, en sentido estricto, fue el impacto del coche lo que causó la muerte del hombre, y fue la acción de hender el puñal lo que causó la muerte de Sara.

Esto sugiere la segunda opción: que las causas y los efectos no sean objetos sino eventos (esto es, sucesos o cambios en los objetos). Donald Davidson es el filósofo moderno más estrechamente asociado a esta concepción[1]. Según ella, cualquier aseveración causal singular que no conecte explícitamente eventos puede ser siempre reformulada de manera tal que conecte eventos, lo que revelaría la verdadera estructura lógica de la aseveración causal. Así, «el coche mató al hombre» debería ser reformulada como:

(a) el impacto del coche causó la muerte del hombre,

y ello pondría de manifiesto que los correlatos son eventos.

Finalmente, de acuerdo con la última de las tres opciones, los correlatos causales serían los hechos. Mientras que es típico de los eventos el que sean identificados por medio de descriptores («la muerte del hombre»), los hechos tienden a ser denotados por medio de oraciones completas («el hombre murió»). De acuerdo con esta opción, defendida recientemente por Hugh Mellor, cualquier oración causal que no conecte explícitamente eventos puede y debe ser reformulada en términos de hechos[2]. Así, «el coche mató al hombre» debería ser vertida como:

(b) el hombre murió porque el coche se movió,

y con ello se daría a entender que los correlatos son hechos.

Si dejamos de lado la primera opción, ¿cuál de las dos restantes sería la correcta? Un argumento habla a favor de la concepción que apela a los eventos y en contra de la que apela a los hechos. Los eventos son cambios en objetos concretos y, por lo tanto, el tipo de cosas verdaderamente aptas para desempeñar un papel causal; los hechos, en cambio, parecen ser abstractos e inertes. Otro argumento, empero, habla a favor de la concepción de los hechos y en contra de la concepción de los eventos: las ausencias pueden ser causas, y las ausencias no son eventos. Mellor propone el ejemplo siguiente: puede ser verdadero decir de Donald que cuando escalaba una montaña «no murió porque no se cayó»[3]. Esto es, el hecho de que Donald no muriese fue causado por el hecho de que no se cayó. Suponiendo, pues, que haya hechos negativos, pero no eventos negativos, este ejemplo hablaría a favor de la concepción de los hechos.

En este capítulo, nuestros ejemplos serán presentados en términos de causalidad de los eventos, pero ello no debería ser entendido como evidencia de que favorecemos esta última opción.

(3) Las relaciones pueden ser clasificadas como reflexivas, simétricas o transitivas[4]. ¿Cuáles de estas dimensiones lógicas atañen a la relación causal? No cabe duda de que la causalidad no es reflexiva: no es cierto que todo evento se cause a sí mismo. ¿Es, por lo tanto, irreflexiva? Es decir, ¿es imposible que un evento se cause a sí mismo? No si los **circuitos causales** [*causal loops*] son lógicamente posibles. Pues, de serlo (lo que podría suceder si fuese posible viajar al pasado), habría casos en los que A causa B, y B causa A; por lo tanto, A se causaría a sí mismo. (Véase el capítulo 6.)

¿Es la causalidad simétrica o asimétrica? Ciertamente no es simétrica. Esto es, no es el caso que si A causa B, B cause A. ¿Es entonces asimétrica? Esto es, suponiendo que A cause B, ¿se sigue acaso que B no causa A? Al menos en nuestro mundo, si A causa B, B no causa A. Pero si los circuitos causales son posibles, habría casos en los que A causaría B y B causaría A. Dada esa posibilidad, la causalidad no sería ni simétrica ni asimétrica: sería tan sólo no simétrica.

¿Es la causalidad transitiva? Esto es, si A causa B, y B causa C, ¿se sigue de ello que A causa C? También en este punto divergen las opiniones. He aquí dos reflexiones que hablan contra la transitividad. Uno, lo más probable es que yo me resista a la idea de que el Big Bang fue lo que hizo que yo escribiera estas líneas; y, sin embargo, eso es precisamente lo que tendría que reconocer si concibiera la causalidad como algo transitivo. Y dos, considérese el ejemplo siguiente. Un alpinista que ve que una roca cae desde lo alto de una montaña se agacha y salva su vida; pero normalmente no diríamos que fue la caída de la roca lo que salvó su vida[5]. Por razones como éstas quizás no deberíamos considerar la causalidad como transitiva.

(4) ¿Qué dirección temporal tiene la causalidad? En nuestro mundo, si A causa B, A es anterior en el tiempo a B. Pero ¿es esto una verdad necesaria? Como veremos, Hume incluyó en su definición de «causa» el que la causa precediera a su efecto. Pero muchos han creído que esto transforma una cuestión sustantiva en un asunto de mera estipulación. Para muchos filósofos, la causalidad retroactiva es teóricamente posible, y cualquier teoría de la causalidad que diga lo contrario es digna de sospecha. De nuevo, nos hallamos en terrenos movedizos, pero parece plausible sostener que, apartando otras variables, es un punto en contra de una teoría de la causalidad el que ésta excluya por definición la posibilidad de una causalidad retroactiva o incluso simultánea.

(5) No tenemos reparo en hablar de *la* causa de un evento. Por ejemplo, en decir que la causa de una explosión fue el que yo encendiera un fósforo. Pero hubo muchas condiciones de fondo [*background conditions*] que tuvieron que estar presentes para que dicho proceso causal se diera (por ejemplo, la presencia de oxígeno, de un fósforo seco, de gas inflamable, etc.). ¿Por qué ninguna de ellas fue seleccionada como «la» causa de la explosión?

Una respuesta famosa es la que dio J. S. Mill (1806-1873). Pensaba que no había distinción alguna entre lo que denominamos «la» causa y otras con-

diciones de la situación respectiva; de hecho, consideró la distinción misma como «arbitraria»[6]. Más recientemente, David Lewis (1942-2002) ha defendido una tesis similar:

A veces destacamos una entre las demás causas de un evento y la designamos «la» causa, como si no hubiera otras; o escogemos a un grupo de ellas como las «causas», y caracterizamos al resto como «factores causales» o «condiciones causales»; o hablamos de la causa «determinante», «real» o «principal». Puede que elijamos causas anormales o extraordinarias, o aquellas que están bajo el control humano, o las que nos parecen buenas o malas, o simplemente aquellas de las que preferimos hablar. No tengo nada que decir acerca de los principios que se emplean en una discriminación tan insidiosa[7].

«Arbitraria» e «insidiosa» son palabras mayores. Sin embargo, en la discusión ulterior se ha revelado muy difícil justificar nuestro modo ordinario de hablar de «la» causa de algún evento. Quizás sea mejor ver el mundo como si estuviera atravesado por procesos causales de los que sólo algunos tienen interés para nosotros. El seleccionar uno de ellos como «la» causa reflejaría meramente nuestros intereses —por ejemplo, lo que consideramos como principal desde un punto de vista explicativo—, y no una distinción real entre causas y «meras» condiciones de posibilidad.

(6) ¿Qué conexión existe entre la causalidad y las leyes de la naturaleza? Algunos creen que cualquier enunciado causal verdadero es sencillamente una instancia de una ley natural. El cemento de la causalidad sería lo que distingue a las leyes de la naturaleza (que tienen la forma «todos los hombres son mortales») de generalizaciones sólo accidentalmente verdaderas (como «todos los hombres llevan camisas»). Sin embargo, como veremos, está lejos de ser obvio que cualquier transacción causal sea *ipso facto* una instancia de una ley natural. No por ello deja de ser plausible afirmar que la necesidad que acompaña a una ley de la naturaleza es la necesidad propia de la causalidad.

Pasemos ahora a la pregunta fundamental: ¿En qué consiste la relación causal?

## La pregunta constitutiva: ¿Qué es la causalidad?

Considérense cuatro eventos, A, B, C y D, de los que A precede a B, y C a D. Supongamos que llegamos a la conclusión de que A causó B, y C causó D, pero que no creemos que A haya causado C, ni que B haya causado D. ¿Qué distinguiría a los pares {{A, B} {C, D}} de los pares {{A, C} {B, D}}? El primer elemento de cada uno de estos pares ocurrió antes que el segundo elemento, y sin embargo hablamos de causa sólo en relación con {A, B} y {C, D}. ¿Qué tendrían, pues, los pares {{A, B} {C, D}} que está ausente en los pares {{A, C} {B, D}}? En otras palabras, si A causa B, ¿qué es lo que *haría verdadero* que A cause B? ¿En virtud de qué, A causa B?

Quienes formulan preguntas semejantes suelen dar por sentado que éstas pueden recibir una respuesta informativa que no contenga ya la idea de causalidad. La tarea del filósofo sería entonces la de analizar las posibles respuestas y seleccionar la más plausible. Esta suposición es razonable como punto de partida. Si nuestro uso de un término es discriminatorio (en el sentido de que se aplica a algunos pares de eventos pero no a otros), entonces resulta plausible creer que hay algo informativo que decir acerca de lo que subyace a dicha discriminación.

Así pues, lo que buscamos son las **condiciones necesarias y suficientes** de la verdad de oraciones de la forma «A causó B». Estamos buscando, pues, algo que tenga la forma: A causó B **si, y sólo si**, X, Y y Z (donde X, Y y Z no presuponen la noción de causalidad). Esto es, siempre que A cause B, también X, Y y Z estarán presentes; y siempre que X, Y y Z estén presentes, A causará B. Como se dijo, una premisa de este experimento es que X, Y y Z puedan ser comprendidos sin necesidad de recurrir a la noción de causalidad. Pero esta premisa podría resultar falsa. Quizás no haya X, Y o Z que hallar.

### La causalidad: ¿Reducible o básica?

Si una propiedad o una relación son consideradas problemáticas, los filósofos a menudo intentan domesticarlas reduciéndolas a componentes más básicos y menos cuestionados. Así, por ejemplo, durante mucho tiempo se consideró que el conocimiento era reducible a creencias verdaderas justificables. Muchos han creído que la causalidad es no menos reducible, ya sea a regularidades de la naturaleza, ya a condicionales contrafácticos. Según la teoría regularista, A causa B si, y sólo si, los eventos del tipo A y los del tipo B aparecen juntos regularmente. Según la teoría contrafactualista, A causa B si, y sólo si, de no haber sucedido A, tampoco lo habría hecho B. Desafortunadamente, ambas teorías están expuestas a contraejemplos, y por lo tanto ninguna de ellas es sostenible. Pero hay otra posibilidad. Quizás la relación causal sea primitiva, y por lo tanto irreducible a elementos más básicos. De acuerdo con esta concepción, nuestro mundo está lleno de necesidad causal. Esta necesidad explica y subyace a las regularidades que observamos a nuestro alrededor, pero no es ella misma reducible a dichas regularidades, ni a ninguna otra cosa.

## Hume

Las discusiones modernas sobre la causalidad tienen su punto de partida en David Hume. Sin embargo, Hume estaba interesado no sólo en la causalidad, sino también en el origen de la idea de causalidad en nuestras mentes. Esta última cuestión era primordial para él, dada la adhesión de Hume al principio empirista siguiente:

(EP) Cada idea es o simple o compleja. Las ideas complejas son construcciones a partir de las ideas simples. Las ideas simples se derivan de las impresiones sensoriales (o de la reflexión).

En el *Tratado sobre la naturaleza humana,* Hume afirma que la idea de causalidad involucra las siguientes relaciones entre objetos (o, como diremos nosotros aquí, entre eventos):

(i) Contigüidad: las causas y los efectos son contiguos en el espacio y en el tiempo.

Si un ladrillo golpea una ventana y la hace añicos, los dos eventos —el golpear del ladrillo en la ventana y el hacerse añicos de ésta— son adyacentes en el espacio y el tiempo. Por supuesto, también decimos, como no deja de observar Hume, que es mi acto de lanzar el ladrillo lo que causa la rotura de la ventana, a pesar de que estos dos eventos no son contiguos. Pero afirma que los dos eventos estarían vinculados entre sí por cadenas intermedias de causa y efecto, cuyos eslabones sí serían elementos contiguos.

(ii) Prioridad: una causa debe preceder temporalmente a su efecto.

Hume considera por un momento la posibilidad de que el término «precede» sea sustituido por la expresión más débil «precede o es simultáneo a»; pero rechaza la propuesta basándose en las razones siguientes: «si cualquier causa pudiera ser perfectamente contemporánea de su efecto, evidentemente [...] todas deberían serlo»[8]. Esto conduciría luego, añade, a «la destrucción de esa sucesión causal que observamos en el mundo; e incluso a la aniquilación misma del tiempo»[9]. Sin embargo, este argumento no es convincente. ¿Por qué habría de seguirse de la afirmación de que algunas causas son simultáneas a sus efectos el que todas deban serlo?

Acaso se piense que Hume está sólo aseverando que (i) e (ii) son típicos rasgos de la causalidad en casos cotidianos normales. Pero está claro que los considera como esenciales a la causalidad de modo absolutamente general. La afirmación de que la contigüidad y la prioridad son esenciales para todas las instancias actuales y posibles de la causalidad dista de ser trivial. Cualquiera que crea en la realidad (o incluso posibilidad) de la acción a cierta distancia temporal o espacial estará negando (i); y cualquiera que crea en la realidad (o incluso posibilidad) de una causalidad retroactiva estará negando (ii). (Véase el capítulo 6.)

Hume acepta (i) e (ii) como esenciales a la causalidad, pero concede que no podemos definir la causalidad únicamente en términos de estos últimos. Y sin duda está en lo cierto. Hay muchos pares de eventos, A y B, donde A y B son contiguos, y A es anterior a B, pero en los que está claro que A no causa B. (Por ejemplo, supóngase que mi perro ladra cada vez que yo le lanzo un ladrillo a una ventana. Las ondas sonoras de su ladrido son anteriores y contiguas

al hacerse añicos de la ventana, y sin embargo no son lo que hace que ésta se vuelva añicos.) ¿Qué más se requiere? Hume sugiere lo siguiente:

(iii) Necesidad: si A causa B, hay una conexión necesaria entre A y B.

Ésta es una condición intuitivamente plausible de la causalidad. Si yo lanzo un ladrillo y éste golpea una ventana, no pensamos que sea una mera casualidad que la ventana se rompa; pensamos que la ventana *debía* romperse. Causa y efecto están conectados necesariamente. Pero ¿qué tipo de necesidad es ésta? No se trata, como Hume nos recuerda, de una necesidad lógica. No hay contradicción lógica alguna en la eventualidad de que un ladrillo simplemente rebote en el vidrio de una ventana. Puede que esto nunca ocurra, pero no es lógicamente imposible. (Tal resultado es concebible, a diferencia de lo que sucede, por ejemplo, con círculos cuadrados o solteros casados, que no son concebibles. Sabemos a priori que jamás nos toparemos con entidades de este último tipo.) La necesidad en cuestión es una necesidad causal o natural, y quedaría por saber si es susceptible o no de análisis ulterior.

Como Hume suscribe (EP), se enfrenta al problema siguiente: ¿Cómo adquirimos la impresión de causalidad, que da lugar a la idea de necesidad causal? En el *Tratado*, Hume afirma que cuando se observan instancias particulares de causa y efecto, todo lo que se percibe es la contigüidad y la sucesión de eventos. No se experimenta la impresión de conexión necesaria entre éstos.

Hume deja el tema así en el *Tratado*, pero lo retoma en la *Investigación*, donde hacer la siguiente descripción de cómo llegamos a la idea de necesidad causal. No podemos derivar la idea de potencia, o conexión necesaria [*power or necessary connection*], partiendo de la observación de una única instancia en la que un evento va seguido de un evento de otro tipo. «Todos los eventos parecen estar enteramente desligados y separados entre sí. Un evento sucede a otro; pero jamás podemos observar una conexión entre ellos. Aparecen *conjuntamente* [*conjoined*], pero nunca *en conexión* [*connected*][10]. Pero a partir de la observación de instancias múltiples podemos adquirir la idea de conexión necesaria: «Cuando una especie particular de evento siempre ha estado, en todas las ocasiones, conectada con otra, ya no tenemos escrúpulo alguno en predecir la ocurrencia de uno en cuanto aparece el otro»[11]. En otras palabras, una vez que hemos presenciado cómo muchos casos de lanzamiento de ladrillo son seguidos de roturas de ventanas, y habiendo constatado que no hay contraejemplos, comenzamos de manera natural a esperar que los futuros lanzamientos de ladrillo vayan seguidos de roturas de ventana. «Esta conexión, por lo tanto, que sentimos en nuestra mente, esta transición habitual desde la representación de un objeto hacia su acompañante usual, es el sentimiento o la impresión a partir de la cual nos formamos la idea de potencia, o conexión necesaria.»[12]

De acuerdo con Hume, la idea de conexión necesaria es una copia de una impresión en nuestra mente, y no la copia de algún rasgo del mundo exterior. Esto es, la exposición repetida a tipos de eventos que aparecen juntos en el mundo hace

que nos formemos naturalmente determinados sentimientos expectantes y anticipatorios, y es de esos sentimientos de donde se deriva nuestra idea de conexión necesaria. Esa idea no es copia de una relación que exista en el mundo. Hume , sin postular conexiones necesarias entre eventos del mundo exterior, nos ofrece una explicación de cómo llegamos a ella. Nuestra creencia natural de que en el mundo existe la necesidad es una **proyección** desde nuestra mente hacia el mundo. De ahí, probablemente, que Hume no haga referencia en la *Investigación* al concepto de necesidad, y en lugar de ello proponga una teoría de la causalidad como regularidad. Si no hay necesidad natural en el mundo, lo que hace verdaderas nuestras aseveraciones causales no puede contener dicha necesidad.

### El proyectivismo de Hume

Para Hume, las ideas son copias de las impresiones, y la idea de causalidad implica la de necesidad. ¿De qué impresión se deriva la idea de necesidad? No de una única exposición a un evento de un tipo seguido por un evento de otro tipo. Sólo la exposición repetida a tales conjunciones de eventos permite que la mente se forme una idea de conexión necesaria. La impresión de necesidad debe, por lo tanto, ser una impresión interna (que consiste en sentimientos de expectativa y anticipación), producida por la exposición de la mente a conjunciones constantes. Por error proyectamos esta necesidad de vuelta al mundo, dando lugar a la ilusión de que hay conexiones necesarias entre eventos del mundo exterior. Tal es la interpretación convencional de lo que Hume pensaba sobre la causalidad, y que hace de él un proyectivista en lugar de un realista sobre la causalidad. Sin embargo, recientemente, algunos filósofos (como Galen Strawson) han cuestionado esta interpretación. ¿Fue Hume un teórico regularista? ¿Rechazó realmente el realismo sobre la causalidad?

## La teoría regularista

Hume escribe: «Podemos definir una causa como *un objeto que es seguido por otro, donde todos los objetos similares al primero son seguidos por objetos similares al segundo.* O, en otras palabras, *donde, de no haber estado presente el primer objeto, el segundo jamás habría existido*»[13].

Como muchos han notado, la frase de Hume «en otras palabras» es sorprendente, ya que la segunda definición en cursivas es completamente distinta de la primera. Tanto es así, que ha dado lugar a una aproximación a la causalidad radicalmente nueva, la teoría contrafactualista, a la que nos referiremos pronto. Por ahora nos interesa su primera definición. Podemos desglosar la propuesta de Hume en los términos siguientes:

(RT) A causó B si y sólo si A precedió a B, y todos los eventos del tipo A son seguidos por eventos del tipo B.

Dicho de otro modo, tras haber observado patrones regulares en la naturaleza, o conjunciones constantes, los clasificamos como interacciones causales. Pero ¿es plausible la teoría de la causalidad como regularidad?

Nótese que la teoría de la causalidad como regularidad pretende ser una explicación reduccionista de la causalidad. Lo que en su caso sigue al «si y sólo si» no hace referencia alguna a la causalidad. Pretende decirnos en términos más familiares qué es la causalidad. Hay, por lo tanto, dos vías expeditas para atacar la teoría regularista. Ésta podría ser verdadera, pero no alcanzar a ser reduccionista, o podría ser falsa. Y podría ser falsa de dos formas. En primer lugar, podría darse el caso de que A causara B, pero o bien A no precediera a B o bien no todos los eventos del tipo A fueran seguidos por eventos del tipo B. En segundo lugar, podría darse el caso de que A precediera a B y que todos los eventos del tipo A fueran seguidos por eventos del tipo B, pero que A no causara B. La teoría de la causalidad como regularidad es argumentativamente vulnerable desde todos estos flancos.

En primer lugar, hay una forma trivial en que la teoría regularista puede fallar de derecha a izquierda. Esto es, habría casos en los que A precede a B y todos los eventos del tipo A son seguidos por eventos del tipo B, y en los que, sin embargo, A no causaría B. Supóngase que A precede a B pero no lo causa, y que A es un evento completamente único: el universo no contiene otros eventos del tipo A. Se trata, obviamente, de una situación posible. Pero de la teoría regularista se sigue que es imposible. En efecto, si A es seguido por B, y A es el único evento del tipo A en el universo, entonces se sigue (trivialmente) que todos los eventos del tipo A son seguidos por eventos del tipo B. En cuyo caso, de acuerdo con la teoría regularista, la condición de la causalidad habría sido satisfecha, y, por lo tanto, A habría causado B. Pero partíamos del supuesto de que A no causó B. Por lo tanto, la teoría regularista, tal como ha sido formulada, tiene que ser falsa.

En segundo lugar, incluso si hay muchos eventos del tipo A, existen escenarios en los que todos los eventos del tipo A son seguidos por eventos del tipo B, pero en los que no se da causalidad alguna. Hace tres siglos, Thomas Reid (1710-1796) señaló que «se sigue de la definición [humeana] de causa que la noche es la causa del día, y que el día es la causa de la noche. Pues desde el comienzo del mundo no hay dos cosas que se hayan sucedido tan constantemente como la noche y el día»[14]. Esta conclusión es, por supuesto, errónea: el día no es causa de la noche, ni viceversa; más bien, día y noche son dos efectos de una causa común (la rotación de la Tierra sobre su eje).

¿Podría fallar la teoría regularista en la dirección contraria? Esto es, ¿podría haber casos en los que A causara B y, sin embargo, o (i) A no precediera a B, o (ii) no todos los eventos del tipo A fueran seguidos por eventos del tipo B? Cualquiera que considere la noción de causalidad retroactiva como coherente sostendrá que podría haber casos en los que A causara B y sin embargo B fuera anterior a A. En cuanto a (ii), ha habido filósofos de la causalidad, denominados «singularistas» (tales como Elizabeth Anscombe y Curt Ducasse), que piensan que del hecho de que A cause B no se sigue que tenga que haber

generalizaciones verdaderas, y libres de excepciones, por las que todos los eventos del tipo A hayan de ser seguidos de eventos del tipo B[15].

Anscombe ofrece cuatro razones a favor del singularismo. Primero, en nuestro discurso etiológico no habría nada que nos obligase a aceptar el universalismo ínsito a la teoría regularista (la concepción de que toda interacción causal está respaldada por una generalización sin excepciones o absoluta). Por ejemplo, uno puede entender perfectamente bien la afirmación de que alguien se contagió de una enfermedad por contacto con un individuo infectado, *sin* que por ello tenga que tomar posición sobre si esta transacción causal está o no respaldada por alguna generalización absoluta.

En segundo lugar, Anscombe apela a desarrollos de la física en el siglo xx. Si los fenómenos cuánticos son **indeterministas** y, sin embargo, hay transacciones causales a nivel cuántico, entonces hay pares de causa y efecto que no caen bajo generalizaciones sin excepciones de la variante «todos los A son B».

En tercer lugar, Anscombe, a diferencia de Hume, piensa que la causalidad es directamente observable en instancias particulares. Si se admite que percibimos directamente cuerpos materiales (y no simplemente fugaces datos de los sentidos), «entonces, ¿qué derecho tiene una teoría de la percepción a negar que percibimos una buena dosis de causalidad?»[16]. Es obvio que esta tesis de observabilidad es perfectamente compatible con el singularismo. Si puedo percibir directamente la causalidad en una instancia particular, lo que suceda en otras partes apenas será relevante para determinar si esa secuencia de eventos es causal o no.

En cuarto lugar, el universalismo tiene la consecuencia extraña siguiente: el que una transacción aparentemente causal sea genuinamente causal dependerá del hecho de si, en algún momento del pasado o del futuro, el lanzamiento similar de un ladrillo fue o será seguido de una rotura similar de la ventana. Esto es contraintuitivo. Cuando decimos que el lanzamiento del ladrillo provocó la rotura de la ventana no estamos haciendo una predicción que se extiende a la totalidad del espacio y el tiempo. El singularismo evitaría ese problema.

La teoría regularista pretende ser una explicación reduccionista de la causalidad. Trata de explicar la causalidad en otros términos —los de precedencia y regularidad—, ninguno de los cuales está presupuesto en, ni presupone, la causalidad. En vista de ello, ni la idea de que A ocurre antes que B ni la de que todos los eventos del tipo A preceden a los eventos del tipo B tendrían por qué presuponer la idea de causalidad. Pero las cosas no son tan sencillas.

¿Qué significa que un evento sea un evento de un tipo particular? Es natural entender la noción, como lo hace Hume, en términos de singularidad. Como A y B son eventos particulares, son irrepetibles. Una determinada rotura de ventana (por ejemplo, la que ocurrió el primero de enero de 2009) no puede volver a ocurrir. Pero eventos similares (eventos de ese mismo tipo) sí podrían hacerlo.

Ahora bien, la similitud puede significar «similitud en todos los respectos» o «similitud en ciertos respectos». El primer tipo de similitud es inservible para el propósito presente: no hay dos roturas de ventana que sean simila-

res en *todos* los respectos, tanto intrínsecos como relacionales. Podrían haber sido causados por personas distintas, o referirse a ventanas distintas; e incluso si fueran causados por las mismas personas, esas personas podrían haber cambiado en varios respectos, cuando menos en su edad. Debido a que el mundo está en perpetuo cambio, no hay dos eventos ocurridos en tiempos diferentes que sean exactamente similares en todos los respectos.

Quizás entendamos por similitud, entonces, una «similitud en ciertos respectos», lo cual encaja mejor con el lenguaje ordinario (si digo que dos saltadores de pértiga son similares, no quiero decir que son similares en todos los respectos, sino en alguno de ellos, como en su técnica para saltar). Así pues, un evento del tipo A sería un evento similar a A en algún respecto. Pero ¿en cuál? Podríamos responder: «en el o los respectos relevantes». Pero eso no es una respuesta: ¿Cuáles serían los respectos relevantes? Se corre el riesgo de que los respectos sólo puedan ser calificados como los que son *causalmente* relevantes: aquellos respectos que son relevantes para dar origen a un evento del tipo B. Pero ahora la individuación de los tipos de eventos estaría presuponiendo la noción de causalidad, en contraste con la pretensión reduccionista que caracterizaba a la teoría regularista[17].

### Elizabeth Anscombe (1919-2001)

Gertrude Elizabeth Margaret Anscombe estudió en St Hugh's College, Oxford, y más tarde asistió a las lecciones que daba en Cambridge el filósofo austríaco Ludwig Wittgenstein. Trabó amistad con éste, tradujo sus *Investigaciones filosóficas* (1953) y se convirtió en uno de sus albaceas intelectuales. Más tarde fue nombrada para ocupar la cátedra que Wittgenstein dejó en Cambridge, la cual ocuparía desde 1970 hasta su jubilación en 1986. Anscombe se hizo famosa no sólo por sus traducciones de Wittgenstein, sino sobre todo por sus libros *Intención* (1957) y *Tres filósofos* (1963) (escrito junto con su marido, el filósofo y lógico P. T. Geach), así como por tres volúmenes de sus artículos reunidos, que abarcan temas de la historia de la filosofía, la metafísica, la filosofía de la mente, la ética y la religión. Figura controvertida, se opuso a la entrada de Gran Bretaña en la Segunda Guerra Mundial (basándose en el argumento de que ello exigiría matar deliberadamente a civiles) y combatió denodadamente el control de la natalidad y el aborto.

## La teoría contrafactualista

Al definir su versión de la teoría regularista de la causalidad, Hume había añadido la frase: «O, en otras palabras, *donde, de no haber estado presente el primer objeto, el segundo jamás habría existido*»[18]. Como se anotó arriba, esta frase es misteriosa, ya que la idea que expresa parece diferenciarse bastante de cualquier versión de la teoría regularista. La noción de un contrafáctico es

más sofisticada que la noción de regularidad o conjunción constante, y muy diferente de ella desde el punto de vista lógico. Nos remite a una nueva (y singularista) teoría de la causalidad, la teoría contrafactualista, cuyo principal portavoz es David Lewis[19]. De acuerdo con la teoría contrafactualista:

(CT) A causó B si y sólo si, de no haber sucedido A, tampoco lo habría hecho B.

Para la teoría contrafactualista, son indispensables los condicionales contrafácticos. Todos nosotros usamos los condicionales contrafácticos de manera constante y a menudo inconsciente. Todos entendemos el significado de «si no te hubieras puesto el cinturón de seguridad, te habrías matado». Llevabas puesto el cinturón de seguridad (por lo que el **antecedente** del condicional es contrario a los hechos), y no te mataste; pero si no te hubieras puesto el cinturón, te habrías matado. La cuestión técnica de cómo analizar este tipo de condicionales es complicada, pero no necesitamos ocuparnos de ella ahora. Todo lo que se nos pide es que apelemos a nuestra forma corriente de entender el mundo y a nuestra comprensión del castellano.

Vale la pena resaltar dos puntos que hablan a favor de la teoría contrafactualista. En primer lugar, en los casos de causalidad más ordinarios, el contrafáctico especificado por la teoría contrafactualista es verdadero. Si lanzo un ladrillo a una ventana, haciendo que ésta se rompa, normalmente será verdadero que, de no haber arrojado yo el ladrillo, la ventana no se habría roto. En segundo lugar, la teoría contrafactualista intenta dar cuenta de la tesis de Hume según la cual la idea de causalidad implica la idea de necesidad. La teoría contrafactualista asume el sentido por el que una causa es necesaria para su efecto: sin la causa, no se habría producido el efecto.

Así pues, ¿deberíamos aceptar la teoría contrafactualista? En absoluto. Admite contraejemplos. En especial, puede haber casos en los que A cause B, pero no sea verdadero que si A no hubiera ocurrido tampoco lo habría hecho B. Los denominados casos de prelación [*pre-emption cases*] son un ejemplo de ello. Supóngase que María ha sido hipnotizada para que se arroje por la ventana en caso de que suene su teléfono a las 8 a.m. Guillermo, al tanto de la situación, la llama a las 8 a.m., provocando que se lance por la ventana. Federico, que también odia a María, estaba entretanto al acecho, para llamar a María en caso de que Guillermo no lo hiciese (pero no se vio obligado a intervenir). Aquí la llamada de Guillermo a María fue lo que causó que ésta saltara por la ventana. Pero si Guillermo no la hubiese llamado, se habría producido exactamente la misma secuencia de eventos (ya que, en ese caso, Federico habría llamado en su lugar). En conclusión, no es verdadero que si Guillermo no hubiera llamado, María habría dejado de arrojarse por la ventana.

Otro tipo de caso es el de la sobredeterminación [*overdetermination*]. Supóngase que A y B causan mancomunadamente C, y que cualquiera de ellos, de hallarse ausente el otro, habría causado C. (Supóngase que María tiene dos

líneas telefónicas, y que el timbre de cualquiera de ellas habría hecho que se lanzase por la ventana. Tanto Guillermo como Federico la llaman a las 8 a.m., haciendo que salte.) C no depende contrafácticamente de A (o de B). Aunque A no hubiera ocurrido, C habría tenido lugar; y aunque B no hubiera ocurrido, C habría tenido lugar. En consecuencia, de acuerdo con la teoría contrafactualista, A no sería lo que causó C, ni B lo que causó C. Pero si ni A ni B causaron C, ¿cómo es que ocurrió C? Un defensor de la teoría contrafactualista podría alegar que fue la conjunción de A y B lo que causó C. Y es cierto que si no hubieran ocurrido ni A ni B, tampoco habría ocurrido C. Pero, en casos en que el resultado está sobredeterminado, ¿cómo es que A y B pueden juntos causar C sin que A cause C y B cause C?[20]

## Una teoría más sencilla

El fallo de la teoría regularista se parece mucho al de la teoría contrafactualista. Ambas se equivocan por no reconocer la función que la necesidad desempeña en la causalidad. Las regularidades se pueden producir accidentalmente, y A puede exigir necesariamente a B aunque B no sea contrafácticamente dependiente de A. Ambas teorías se aferran a contingencias que resultan verdaderas en la mayoría de los casos de causalidad, pero que no corresponden a su misma esencia. Las causas y los efectos exhiben normalmente un patrón regular, y los efectos dependen contrafácticamente, en la generalidad de los casos, de sus causas. Pero ninguna de las dos teorías capta plenamente nuestra idea de causalidad.

Puesto que ninguna de las dos es satisfactoria, quizás deberíamos considerar una teoría de la causalidad más simple y no reductiva, la teoría sencilla:

(ST) A causa B si y sólo si A exige necesariamente (junto con otras condiciones circunstanciales) la ocurrencia de B,

donde la necesidad involucrada es una necesidad causal, una relación que se asume como primitiva. La necesidad causal debería ser distinguida de la necesidad lógica (esto es, el tipo de necesidad exhibida por «o está lloviendo o no lo está»). Como ya observara Hume, si A causa B, nunca es lógicamente contradictorio suponer «A pero no B». Esta especie no lógica de necesidad es considerada un ingrediente básico de nuestra ontología. A diferencia de la teoría regularista y de la teoría contrafactualista, la teoría sencilla ofrece una explicación no reduccionista de la causalidad: la noción de «causa» y la noción relacionada de «necesidad causal» no son susceptibles de un análisis ulterior. No podemos explicar la idea de necesidad causal en términos diferentes y más básicos.

Una de las objeciones a la teoría sencilla es de índole epistémica. Si todo lo que observamos son las regularidades, ¿por qué hemos de creer en la nece-

sidad causal? Siguiendo a Anscombe, podríamos replicar que la causalidad puede ser observada directamente en ciertos casos, como cuando decidimos mover nuestros brazos, o cuando vemos cómo un cuchillo corta la mantequilla. En segundo lugar, podemos apelar en otros casos de causalidad al principio de «inferencia de la explicación mejor». Esto es, postular la necesidad causal es lo que mejor explicaría las regularidades que observamos en nuestro entorno. ¿Por qué se rompe la ventana cada vez que le arrojo un ladrillo? Porque es causalmente necesario que una ventana se rompa cada vez que se le arroja un ladrillo. La necesidad causal sería algo que postulamos para explicar el mundo que nos rodea. Sería un postulado teórico de sentido común. Ciertamente disponemos de buena evidencia de causalidad cuando disponemos de regularidades y dependencias contrafácticas. Pero éstas no constituyen la esencia de la causalidad. Son indicios de algo que subyace a ellas: la necesidad causal.

## Observaciones finales

Hemos constatado que dos teorías de la causalidad deliberadamente reduccionistas, la teoría regularista y la teoría contrafactualista, están expuestas a objeciones. Una alternativa natural es la teoría sencilla de la causalidad, que posee los rasgos siguientes:

- Sostiene que la causalidad implica necesidad (causal o natural).
- Afirma que la noción de necesidad causal no admite un análisis reduccionista.
- Encaja bien con la concepción singularista de la causalidad.
- Se ajusta tanto a la concepción de que la causalidad es a veces directamente observable como a un postulado teórico que explica las regularidades que observamos a nuestro alrededor.
- Finalmente, la teoría sencilla nos permite distinguir las leyes de la naturaleza de las generalizaciones accidentalmente verdaderas, mediante el hecho de que las leyes naturales son verdaderas por necesidad (natural).

La teoría simple es, pues, digna de ser tomada en serio.

## Preguntas para el estudio

- ¿Cuáles son los correlatos de la relación causal?
- ¿Cuán plausible es el principio empirista de Hume (EP)?
- ¿Es posible la causalidad retroactiva?
- ¿Cuál es la objeción más contundente contra la teoría regularista de la causalidad?
- ¿Por qué los casos de prelación y sobredeterminación son problemáticos para la teoría contrafactualista?

## Notas

1. Véase, por ejemplo, «Causal Relations», reimpreso en sus *Essays on Actions and Events* (Oxford: Oxford University Press) 1985, pp. 149-162.
2. D. H. Mellor, *The Facts of Causation* (Londres: Routledge) 1995.
3. Ibíd., p. 132.
4. Una relación R es reflexiva exactamente si necesariamente, para cualquier objeto *a*, *a*R*a*. Una relación es simétrica exactamente si necesariamente, para cualesquiera objetos *a* y *b,* si *a*R*b,* entonces *b*R*a.* Una relación es transitiva exactamente si necesariamente, para cualesquiera objetos *a, b* y *c*, si *a*R*b* y *b*R*c,* entonces *a*R*c.* La relación «tan alto como», por ejemplo, es reflexiva, simétrica y transitiva. Necesariamente: todo objeto es tan alto como sí mismo; si *a* es tan alto como *b, b* es tan alto como *a;* y si *a* es tan alto como *b,* y *b* es tan alto como *c*, entonces *a* es tan alto como *c*. La relación «ama» no es, por desgracia, reflexiva, simétrica ni transitiva.
5. El ejemplo se debe a Ned Hall. Véase «Causation and the Price of Transitivity», *Journal of Philosophy,* vol. 97, 2000, pp. 198-222.
6. J. S. Mill, *A System of Logic* (Nueva York: Harper & Brothers) 1846, p. 198.
7. D. Lewis, «Causation», reimpreso en sus *Philosophical Papers,* vol. II (Oxford: Oxford University Press) 1986, p. 162. La concepción que Lewis defiende aquí encaja con su teoría contrafactualista de la causalidad: dondequiera que haya una dependencia contrafáctica habrá causalidad.
8. D. Hume, *A Treatise Concerning Human Understanding,* en A. J. Ayer y R. Winch (eds.), *British Empirical Philosophers* (Londres: Routledge) 1965, p. 359.
9. Ibíd., p. 359.
10. D. Hume, *An Enquiry Concerning Human Understanding,* ed. de L. A. Selby-Bigge (Oxford: Oxford University Press) 1962, p. 74.
11. Ibíd., pp. 74-75.
12. Ibíd., p. 75.
13. Ibíd., p. 76.
14. T. Reid, *Essays on the Active Powers of the Human Mind,* ed. de Baruch A. Brody (Cambridge, Mass.: MIT Press) 1969, p. 334.
15. Véanse G. E. M. Anscombe «Causality and Determination» y C. J. Ducasse «On the Nature and the Observability of the Causal Relation», ambos reimpresos en E. Sosa y M. Tooley (eds.), *Causation* (Oxford: Oxford University Press) 1993, pp. 88-104 y pp. 125-136, respectivamente.
16. Anscombe, *op. cit.*, p. 92.
17. Esta objeción es debida a R. Taylor. Véase su *Metaphysics* (Englewood Cliffs, NJ: Prentice Hall) 1992, capítulo 10.
18. Hume, *Enquiry,* p. 76.
19. Véase D. Lewis, *op. cit.*
20. Hay que distinguir la sobredeterminación, por un lado, del caso en que dos personas juntas producen determinado efecto pero el efecto no se habría producido si alguna de las dos no hubiera actuado como lo hizo, por el otro. En este último caso, no es difícil entender cómo D y E actuaron juntos para causar F a pesar de que D no causara F ni E causara F. D no causó F porque D no bastaba, por sí solo, para causar F. Pero esto es justamente lo que no se puede decir en el caso de la sobredeterminación.

## Bibliografía adicional comentada

G. E. M. Anscombe, «Causality and Determination», en E. Sosa y M. Tooley (eds.) *Causation* (Oxford: Oxford University Press) 1993, pp. 88-104. Un contundente alegato antihumeano. Anscombe arguye que las relaciones causales no presuponen generalizaciones absolutas, y que la causalidad puede ser percibida directamente. Artículo difícil.

P. Horwich, «Lewis's Programme», reimpreso en Sosa y Tooley, *op. cit.,* pp. 208-216. Una crítica útil, aunque exigente, de la teoría contrafactualista de Lewis.

D. Lewis, «Causation», reimpreso en Sosa y Tooley, *op. cit.,* pp. 193-204. La fuente clásica de las aproximaciones contrafactualistas a la causalidad. Una lectura ardua.

E. Sosa y M. Tooley, «Introduction», en Sosa y Tooley, *op. cit.,* pp. 1-32. Una introducción amena a los problemas discutidos en este capítulo.

G. Strawson, *The Secret Connexion* (Oxford y Nueva York: Clarendon Press) 1989. Strawson hace una defensa detallada de una interpretación de Hume como realista sobre la causalidad.

## Recursos en Internet

J. Faye, «Backward Causation», *The Stanford Encyclopedia of Philosophy (edición de otoño de 2005),* Edward N. Zalta (ed.). Bajado el 31 de mayo de 2006 de <http://plato.stanford.edu/archives/fall2005/entries/causation-backwards>.

D. Garrett (2005) «Hume, David», en E. Craig (ed.), *Routledge Encyclopedia of Philosophy.* Bajado el 31 de mayo de 2006 de <http://www.rep.routledge.com/article/DB040>.

P. Menzies (2001), «Counterfactual Theories of Causation», *The Stanford Encyclopedia of Philosophy (edición de primavera de 2001),* Edward N. Zalta (ed.). Bajado el 31 de mayo de 2006 de <http://plato.stanford.edu/archives/spr2001/entries/causation-counterfactual>.

J. Schaffer (2003) «The Metaphysics of Causation», *The Stanford Encyclopedia of Philosophy (edición de primavera de 2003),* Edward N. Zalta (ed.). Bajado el 31 de mayo de 2006 de <http://plato.stanford.edu/archives/spr2003/entries/causation-metaphysics>.

# 5. El tiempo: la cuestión fundamental

## Introducción

En este capítulo nos ocuparemos de una pregunta fundamental sobre nuestro mundo: ¿Qué es el tiempo? Para centrar nuestra discusión, examinaremos en detalle algunas ideas de J. M. E. McTaggart (1866-1925). Los escritos de McTaggart determinan buena parte de la agenda de las discusiones filosóficas del siglo XX sobre el tiempo. Aunque la tesis general de McTaggart, con su conclusión de que el tiempo es irreal, es hoy generalmente rechazada, sus dos argumentos subsidiarios tienen todavía partidarios.

### J. M. E. McTaggart (1866-1925)

John McTaggart nació en Wiltshire, Inglaterra. Estudió en el Trinity College, Cambridge, y enseñó en él durante casi toda su vida ( entre sus discípulos estuvo Bertrand Russell). En su artículo más famoso, «La irrealidad del tiempo» (1908), distinguió dos formas de ordenar los eventos en el tiempo, la serie A y la serie B. Argumentó que la serie A es fundamental para el tiempo, pero que desemboca en una contradicción. Infirió de ello que el tiempo era irreal. También sostuvo la irrealidad del espacio. Consideró al mundo espaciotemporal como una ilusión, y creyó que la realidad estaba formada de almas inmateriales, en extática comunión recíproca. Junto con F. C. Bradley (1816-1924), McTaggart fue un miembro notable de la efímera escuela del idealismo británico, influida por Hegel, que empezó a declinar hacia finales del siglo XIX y comienzos del XX.

## Planteamiento del debate: las series A y B

McTaggart comienza su explicación sobre el tiempo observando lo siguiente:

> Las posiciones en el tiempo, la manera como el tiempo se nos aparece *prima facie,* son fijadas de dos maneras. Cada posición es Anterior y Posterior con respecto a algunas de las demás posiciones [...] En segundo lugar, cada posición es o Pasada o Presente o Futura. Las distinciones de la primera clase son permanentes, a diferencia de las segundas. Si M fue alguna vez anterior a N, siempre lo será. Pero un evento que ahora es presente fue futuro, y será pasado[1].

A la primera serie de posiciones McTaggart la denomina la serie B, y a la segunda, la serie A. Es bastante plausible afirmar, como hace McTaggart, que las posiciones en el tiempo son distinguidas por nosotros de las dos maneras mencionadas, y que las posiciones de la serie B son permanentes de una forma en que no lo son las de la serie A. Esto queda de manifiesto en la invariabilidad del valor de verdad de las atribuciones de la serie B. Si es verdad que mi nacimiento fue anterior al año 1970, entonces siempre fue verdad y siempre será verdad que mi nacimiento ocurrió antes del año 1970. El valor de verdad de la oración «mi nacimiento es anterior a 1970» nunca varía. Por contraste, las atribuciones de la serie A son provisionales en lugar de permanentes. Mi nacimiento fue futuro, luego se hizo presente y luego se convirtió cada vez más en pasado. El valor de verdad de la oración «mi nacimiento es futuro» varió, pasando de verdadero a falso.

A pesar de ello, las series A y B no son series temporales independientes entre sí. Hay **nexos veritativos** [*truth-value links*] obvios e innegables entre las atribuciones de la serie A y las de la serie B. Por ejemplo, mi oración en 2006 «la muerte de Hitler ocurrió en el pasado» es verdadera si y sólo si la muerte de Hitler fue anterior a 2006; mi oración en 2006 «la muerte de Blair es futura» es verdadera si y sólo si la muerte de Blair es posterior a 2006; y la oración proferida el 1º de febrero de 2009 «está lloviendo ahora» es verdadera si y sólo si está lloviendo el 1º de febrero de 2009. Estos nexos veritativos podrían inducir a pensar que la distinción entre las series A y B es convencional en lugar de metafísica. Es decir, alguien podría pensar que las series A y B son respecto del tiempo lo mismo que, por ejemplo, la escala Fahrenheit y la escala centígrada son respecto de la temperatura, o las pulgadas y los centímetros respecto de la distancia: meras variantes notacionales. Eso sería, por descontado, una concepción posible, pero no es la de McTaggart. Éste sostiene que una de estas series (la serie A) es más fundamental para el tiempo que la otra, y que, por lo tanto, la teoría B —que afirma que la serie B es esencial al tiempo— es falsa.

## Las series A y B

McTaggart observó que las posiciones en el tiempo pertenecen a dos series, a las que denominó «la serie A» y «la serie B». Cada evento pertenecería a ambas series, y, sin embargo, las series tienen características muy distintas. Los eventos que están en la serie A cambian continuamente de posición. Los eventos futuros se hacen presentes; los presentes se hacen pasados, los pasados se hacen aún más pasados. Por contraste, la posición de un evento en la serie B no cambia nunca. Si un evento ocurrió en 1961, siempre fue y será cierto que ocurrió en 1961. Si un evento es anterior a otro evento, ese hecho tampoco cambia nunca. Se pueden dar tres concepciones con respecto a las series A y B: que la serie A es fundamental para el tiempo; que la serie B es fundamental para el tiempo, y que ninguna de las dos series es fundamental. Un defensor de la primera concepción es un teórico A. Un defensor de la segunda es un teórico B. Un defensor de la tercera concepción piensa que el debate está mal planteado: habría dos lenguajes del tiempo pero una única serie que puede ser descrita de modos diferentes.

## La serie A es fundamental para el tiempo

De acuerdo con McTaggart, el hecho de que las posiciones de la serie B sean permanentes no nos debería conducir a pensar que son más objetivas o más esenciales para el tiempo que las posiciones de la serie A. Este autor piensa, por el contrario, que aunque las determinaciones A y B sean esenciales para el tiempo, las determinaciones de la serie A son más **fundamentales** que las de la serie B.

La distinción entre lo esencial y lo fundamental es, ciertamente, muy sutil. En su uso ordinario, las palabras «fundamental» y «esencial» son intercambiables. ¿Qué distinción, entonces, estaría tratando de mostrarnos McTaggart?

Quizás se piense que quiere afirmar que, a pesar de que tanto la serie A como la serie B son esenciales (o necesarias) para el tiempo, únicamente la serie A basta por sí sola para el tiempo. Quizás éste sería el sentido de afirmar que la serie A es más fundamental que la serie B. Sin embargo, McTaggart concede que la serie B «no puede existir sino dentro de la temporalidad», sugiriendo así la suficiencia de la serie B para constituir el tiempo[2]. Pero, en otro pasaje, dice que la serie B «no es suficiente por sí sola para constituir el tiempo»[3]. ¿Cómo se explica esta aparente disparidad?

Lo que McTaggart podría estar queriendo decir es lo siguiente. Una descripción de la realidad llevada a cabo únicamente en términos de la serie B es necesariamente incompleta. El conjunto de los hechos de la serie B es un subconjunto de la totalidad de los hechos temporales. Esta falta de completitud significaría que la serie B no puede ser fundamental para el tiempo. En contraste, la serie A no adolece de esa falta de completitud.

Y ¿por qué habría de pensar McTaggart que la serie A es fundamental para el tiempo? McTaggart expone primero su argumento y luego examina una objeción que le había sido planteada por Bertrand Russell. Su argumento puede formularse de la manera siguiente:

(1) El tiempo involucra necesariamente el cambio.
(2) El cambio sólo es posible en la serie A.
Por lo tanto: (3) El tiempo involucra fundamentalmente la serie A.

McTaggart no ofrece argumento alguno para la premisa (1). Sólo afirma que «no podría haber tiempo, si no hubiera algo que cambiara». Algunos filósofos, especialmente Shoemaker, han sostenido que es posible el tiempo sin el cambio (véase el capítulo 6). Sin embargo, McTaggart entiende (1) como un truismo, pues cree que el paso del tiempo (o «devenir puro») ya es de por sí un cambio. Pero, entendida de este modo, la premisa (1) da por sentado precisamente aquello que debería probar frente al teórico B. Afortunadamente, sin embargo, McTaggart no necesita realmente la premisa (1). La verdad de la premisa (2) basta para demostrar la superioridad de la serie A sobre la serie B.

Pues bien, McTaggart sí tiene un argumento a favor de la premisa (2). Para sustentarla, apela a la permanencia de las localizaciones y las relaciones de la serie B. Los eventos nunca alteran su localización en la serie B (siempre ha sido y siempre será verdadero que la muerte de Hitler ocurrió en 1945) ni la relación de sus posiciones en la serie B con otros eventos (siempre ha sido y siempre será verdadero que la muerte de Hitler ocurrió después de la muerte de César). Esto, para McTaggart, basta para mostrar que la serie B no admite la posibilidad del cambio.

La única forma en que se puede decir que un evento —tal como la muerte de la reina Ana— está inmerso en el cambio sería la siguiente: «Fue en cierta ocasión un evento en el futuro lejano. Con cada momento que pasaba, se convertía en un evento futuro más próximo. Finalmente, se hizo presente. Luego se convirtió en pasado, y siempre habrá de permanecer en el pasado, aunque, con cada momento que pasa, se encontrará más y más alejado en el pasado»[4]. (Según la concepción de McTaggart, el pasado está cambiando continuamente: los eventos pasados se convierten en cada vez más pasados.) Por todo ello, concluye McTaggart, el cambio sólo es posible en la serie A, y la serie A es fundamental para el tiempo.

## La réplica de Russell

Russell no creía que la serie A fuera fundamental para el tiempo. De acuerdo con él,

pasado, presente y futuro pertenecen al tiempo no *per se,* sino únicamente en relación con un sujeto cognoscente. Una aserción de que N es presente significa que N es simultá-

neo a esa aserción; una aserción de que N es pasado o futuro significa que N es anterior o posterior a esa aserción [...] Si no hubiera consciencia [...] nada sería pasado, presente o futuro[5].

De lo anterior se sigue que la serie A es relativa al sujeto o dependiente de la mente, y, por lo tanto, que no es fundamental. Como Russell afirma que lo fundamental para el tiempo es la serie B, podríamos considerarlo un teórico B. Un teórico A, por contraste, sostiene que lo fundamental para el tiempo es la serie A.

¿Cómo responde entonces Russell al argumento de McTaggart contra la teoría B? Russell y McTaggart difieren sobre cómo caracterizar el cambio. Según la caracterización de Russell, la premisa (2) es falsa. Russell escribe:

El cambio es la diferencia, en cuanto a verdad y a falsedad, entre una proposición que concierne a una entidad y el tiempo T y una proposición que concierne a la misma entidad y el tiempo T*, siempre y cuando estas proposiciones se diferencien únicamente por el hecho de que la una se da en T, mientras que la otra se da en T*[6].

De este modo, hay cambio si, por ejemplo, la proposición «en el tiempo T mi atizador está caliente» es verdadera, y la proposición «en el tiempo T* mi atizador está caliente» es falsa. Dicho de manera más sencilla, hay cambio si mi atizador está caliente en un momento y no está caliente en otro. Y el cambio, así entendido, requeriría únicamente la serie B.

McTaggart tiene una réplica para Russell. Según McTaggart, la explicación de Russell simplemente no es una explicación del cambio. Pues si la proposición «en un tiempo (cualquiera dado) T mi atizador está caliente» es verdadera, es siempre verdadera; y si la proposición «en un tiempo T* mi atizador está caliente» es falsa, es siempre falsa. Y «esto no determina cambio alguno en las cualidades del atizador»[7]. Es siempre verdadero, en todo tiempo, que el atizador está caliente en T, y que no está caliente en T*. Esta permanencia en el valor de verdad implica, para McTaggart, que no hay cambio alguno en el atizador.

McTaggart expresa la misma idea en términos de hechos. El hecho de que el atizador esté caliente en T nunca cambia; así como tampoco cambia el hecho de que no esté caliente en T*. Sin embargo, «no puede haber cambio a menos que cambien los hechos»[8]. Y los únicos hechos que pueden cambiar son los hechos de la serie A. Si T se refiere al presente, el atizador está ahora caliente y estará frío. El hecho de que esté caliente dará paso al hecho de estar frío, cuando este último se haga presente. Por lo tanto, concluye McTaggart, el cambio sólo es posible en la serie A.

¿Cómo podríamos dirimir esta disputa? Se nos ofrecen dos explicaciones del cambio harto diferentes, a las que podríamos denominar «cambio McTaggart» y «cambio Russell». El cambio McTaggart ocurre cada vez que un evento o un hecho alteran su posición en la serie A. El cambio Russell ocurre cada

vez que un objeto tiene, en tiempos diferentes, propiedades incompatibles. ¿Se puede admitir el cambio Russell como una noción de tiempo válida?

*Prima facie,* la respuesta es afirmativa. El que un objeto altere sus propiedades —por ejemplo, el que la valla de mi jardín sea verde el lunes y luego sea pintada de rojo el martes— contaría ordinariamente como cambio. Para que un objeto cambie, no se requiere que el evento de pintar o el hecho de que la valla sea verde cambien ellos mismos. Simplemente se requiere que el objeto haya cambiado de propiedades. En tal caso, la réplica de Russell se sostiene, y la razón que McTaggart aduce a favor de (2) se desmorona.

Sin embargo, hay otra objeción al argumento de McTaggart. Al presuponer que el cambio requiere el cambio de un evento o de un hecho, y dado que la única forma de que un evento o hecho cambie es que lo haga con respecto a su posición en la serie A, McTaggart ha cometido una petición de principio frente a su oponente, el teórico de la serie B. Ningún teórico de la serie B aceptará una definición de «cambio» que simplemente incorpore de entrada una referencia a la serie A. El argumento de McTaggart presupone, por lo tanto, lo que tendría que demostrar.

### El cambio

¿Qué es el cambio? De acuerdo con Russell, el cambio consiste en que un objeto tenga propiedades diferentes e incompatibles en tiempos diferentes. Si la valla de mi jardín es verde el lunes y luego es pintada de rojo el martes, habrá cambiado. De acuerdo con McTaggart, el cambio es algo que le ocurre no a los objetos sino a los eventos. Ningún evento puede modificar su lugar en la serie B: éste se encuentra determinado y es permanente. Un evento cambia únicamente con respecto a su posición en la serie A: fue primero futuro, después fugazmente presente y finalmente cada vez más pasado. McTaggart pensó, por lo tanto, que el cambio sólo era posible en la serie A. Muchos han sostenido que el «cambio Russell» corresponde a nuestra noción ordinaria de cambio. Decimos que un objeto cambia, y el cambiar de un objeto es un evento, un suceso. ¿Acaso debería cambiar el evento para que hubiese cambio? Los teóricos B creen que no. También algunos teóricos A piensan que no. Éstos opinan que el «cambio Russell» es una noción perfectamente aceptable del cambio, pero que la serie A es fundamental para el tiempo por otras razones.

## La paradoja de McTaggart

Uno esperaría que McTaggart se hubiese conformado con establecer la teoría A del tiempo y se quedase en ese punto. Pero eso no fue lo que sucedió. Una vez que se convenció de que el tiempo involucra fundamentalmente la serie A, McTaggart procedió a argumentar que la serie A es contradictoria, y que,

por lo tanto, el tiempo es irreal. McTaggart no tuvo reparos en abrazar la conclusión de que el paso del tiempo es una ilusión y que nada cambia realmente. Aunque esta conclusión difícilmente concuerda con el sentido común, el argumento de McTaggart de que la serie A es contradictoria —conocido como la paradoja de McTaggart— es un argumento interesante e ingenioso en virtud de sí mismo.

McTaggart escribe:

> El pasado, el presente y el futuro son determinaciones incompatibles [...] Pero todo evento las tiene todas. Si M es pasado, ha sido presente y futuro. Si es futuro, será presente y pasado. Si es presente, ha sido futuro y será pasado. Así, las tres características pertenecen a cada evento[9].

Podemos representar el razonamiento de McTaggart del modo siguiente:

(4) Todo evento es pasado, presente y futuro.
(5) Ningún evento puede ser pasado, presente y futuro.
Por ende: (6) La serie A es contradictoria.

La idea que subyace a (4) es que ningún evento escapa al paso del tiempo: cada evento es futuro, luego presente efímero y luego pasado para siempre. Cada evento ocupa cada posición de la serie A. (Habría excepciones, si hubiera un primer evento o un último evento. Pero esto no afectaría al argumento, ya que el primer evento sería hoy presente y pasado, y el último evento, futuro y presente, y estas determinaciones son respectivamente incompatibles entre sí.)

La idea que subyace a (5) es que pasado, presente y futuro son determinaciones incompatibles: si un evento es pasado, no es presente o futuro; si es presente, no es pasado o futuro, y así sucesivamente. Nada puede poseer características incompatibles. De (4) y (5), se sigue (6).

Ahora bien, una obvia réplica a este argumento viene a la mente. Y es el propio McTaggart quien la formula:

> Nunca es cierto, se responderá, que M *sea* presente, pasado y futuro. *Es* presente, *será* pasado y *ha sido* futuro. O bien *es* pasado, y *ha sido* futuro y presente; o *es* futuro y *será* presente y pasado. Las características sólo son incompatibles cuando son simultáneas, y no hay contradicción al respecto por el hecho de que cada término las tenga todas sucesivamente[10].

Así pues, hay un sentido por el que (4) es verdadero, pero por el cual (5) es falso; y hay un sentido por el que (5) es verdadero, pero por el cual (4) es falso. Consecuentemente, el razonamiento (4)-(6) es inválido, ya que tiene premisas incompatibles.

Pero McTaggart tiene preparada una respuesta ingeniosa a esta réplica. Podemos evitar el cargo de contradicción en las tres posiciones del nivel bási-

co de la serie A (pasado, presente y futuro) invocando tres posiciones del segundo nivel de la serie A (por ejemplo, M es presente, fue futuro y será pasado). Pero hay nueve posiciones en esta serie de segundo nivel (es pasado, es presente, es futuro; fue pasado, fue presente, fue futuro; será pasado, será presente, será futuro), y cada evento ocupa cada una de estas posiciones de la serie A. Algunas combinaciones de estas nueve posiciones son incompatibles entre sí (por ejemplo, es presente y es pasado). Podemos evitar estas contradicciones distinguiendo tiempos más complejos y moviéndonos a posiciones del tercer nivel de la serie A. Pero algunas de las veintisiete posiciones resultantes serán incompatibles entre sí. Para evitar la contradicción, deberemos entonces ascender a un cuarto nivel, y así sucesivamente. Podemos siempre sustraernos a la contradicción respectiva ascendiendo un nivel, pero en cualquier nivel se mantendrá alguna contradicción. «Y, como esto continúa indefinidamente, el primer conjunto de términos nunca logra realmente sustraerse a la contradicción.»[11] La «réplica obvia» no resulta, después de todo, tan obvia.

## Diagnóstico

La paradoja de McTaggart, para usar nuestro apelativo, tiene, como Jano, un doble rostro: un día nos parece convincente y, al siguiente, sofística. Pero considero que el teórico A (que la paradoja toma como blanco) tiene una respuesta. En lo que sigue, tomo prestadas algunas ideas de Michael Dummett y Paul Horwich.

Durante el desarrollo de su paradoja, McTaggart hace, al decir de Dummett, una presuposición implícita[12]. Supone que una descripción consistente y completa de la realidad es, en principio, posible. Es decir, que, independientemente de la posición que uno ocupe en el tiempo, es posible dar una descripción de la realidad que sea consistente e incluya todas las verdades. Lo que la paradoja de McTaggart revela es que la teoría A es incompatible con esa presuposición. Cuando tratamos de especificar todas las verdades de la serie A, en lugar de especificar únicamente aquellas que son verdaderas desde la perspectiva que ocupamos actualmente, terminamos contradiciéndonos. De ahí que, si la teoría A es verdadera, sea imposible dar una descripción completa y consistente de la realidad. Como McTaggart presuponía que tal descripción es posible, concluyó que la serie A era contradictoria. Por contraste, una descripción de la realidad en términos de la serie B es consistente, pero incompleta (de acuerdo con McTaggart, deja por fuera el fenómeno del cambio). Así, también la teoría B violaría abiertamente, pero en un sentido diferente, la presuposición de McTaggart.

Ahora bien, ¿tiene el teórico A que aceptar la presuposición de McTaggart de que, en principio, es posible una descripción completa y consistente de la realidad? Es indudable que debería aceptar el requisito de la consistencia. Pero ¿necesita aceptar también el requisito de completitud, esto es, el requisi-

to de que la realidad puede ser completamente descrita, independientemente de la perspectiva temporal que se adopte?

Si se reflexiona un poco, se verá que la respuesta es que no. Es esencial a la teoría A el que los hechos temporales fundamentales (es decir, los hechos temporalmente conjugados que registran la posición de un evento en la serie A) cambien a medida que lo hace el tiempo. De acuerdo con la teoría A, la realidad está siendo constantemente redistribuida: hechos que son futuros se vuelven presentes; hechos que son presentes se vuelven pasados; hechos que son pasados se vuelven aún más pasados. La realidad está cambiando continuamente. Si ahora estamos en el 2009, puedo enunciar los hechos tal como son desde esa perspectiva, incluyendo el hecho de que mi muerte es futura. En el 3009 la formulación de los hechos incluirá el hecho de que mi muerte es pasada. Sólo habría contradicción si presuponemos que debe haber una descripción (sin perspectiva definida) que incorpore ambos hechos. El teórico A negaría que exista semejante descripción: cualquier descripción de la realidad temporal, desde el interior del tiempo, sería necesariamente incompleta. Y eso le permitiría a la teoría A escapar a la paradoja de McTaggart.

Dummett admite que no es fácil renunciar a la creencia de que debe existir una descripción completa de la realidad, que «de cualquier cosa que sea real debe haber una descripción completa, es decir, independiente de cualquier observador»[13]. Sin embargo, está dispuesto a interpretar el razonamiento de McTaggart en el sentido de que «debemos abandonar nuestro prejuicio de que existe necesariamente una descripción completa de la realidad»[14].

Paul Horwich piensa, en cambio, que no se debería renunciar al requisito de la completitud, y concluye, por lo tanto, que la paradoja de McTaggart logra refutar la teoría A del tiempo[15]. ¿Por qué considera Horwich como sacrosanto el requisito de la completitud? Piensa que la idea del teórico A de que hay «una variación, de un tiempo a otro, acerca de qué hechos se dan» se basa en «una concepción idiosincrásica e infundada de *hecho*»[16]. En otras palabras, abandonar el requisito de la completitud (como lo hace el teórico A) violaría nuestra comprensión ordinaria de lo que es un hecho; por lo tanto, dicho requisito debería ser respetado.

Horwich proporciona un ejemplo para sustentar su tesis. Dice que:

no consideramos que

X está a la izquierda de Y

y

X no está a la izquierda de Y

sean formulaciones explícitas de hechos. Antes bien, suponemos que cada vez que tales aseveraciones son verdaderas, se trata de formulaciones parciales de hechos cuyas descripciones explícitas asumen la modalidad siguiente:

X está a la izquierda de Y en relación con Z

y

X no está a la izquierda de Y en relación con W [...].

El punto general sería que reservamos el término «hecho» para aquellos aspectos de la realidad cuyas descripciones explícitas son oraciones que son verdaderas *simpliciter* —y no meramente verdaderas en relación con algunos contextos o puntos de vista, y falsas en relación con otros[17].

¿Cómo debería responder un teórico A? Es significativo que el ejemplo de Horwich sea espacial. Esto, diría el teórico A, es crucial. Pues es plausible suponer que puede darse una descripción completa y consistente de todos los hechos espaciales. La intuición de completitud es robusta en el caso de los hechos espaciales. Al fin y al cabo, ésta es la razón de que los mapas resulten útiles: no necesitamos un mapa diferente para cada lugar. Por lo tanto, Horwich tendría razón en su ejemplo de que la descripción correcta es «X está a la izquierda de Y en relación a Z», en lugar de «X está a la izquierda de Y»; pero ello se debería a que los hechos espaciales admiten una descripción completa (independiente de todo observador). Sin embargo, los hechos temporales no la admiten, y el ejemplo de Horwich serviría únicamente para destacar la diferencia fundamental que existe entre el tiempo y el espacio. O al menos ésa sería una manera de responder del teórico A.

Otra manera sería la siguiente. Aunque uno pudiera formular una «teoría A» del espacio —la concepción de que hay propiedades espaciales fundamentales aquí y allí, próximas y lejanas, no reducibles a propiedades y relaciones espaciales de la «serie B» como «en Oxford», «80 kilómetros al norte de Londres», etc.—, tal teoría no tendría plausibilidad alguna. Como dice Dummett: «el uso de **expresiones de ocurrencia reflexiva** espacial [*spatially token-reflexive expressions*] [es decir, expresiones tales como «aquí» y «allí»] no es esencial a la descripción de objetos como espacialmente localizados. Esto es, puedo describir una disposición de objetos en el espacio aunque yo mismo no ocupe una posición en ese espacio»[18].

En otras palabras, puedo proporcionar una descripción completa de los hechos espaciales desde cualquier perspectiva espacial, o desde ninguna, y, por lo tanto, sin usar términos como «aquí». En cambio, según Dummett, no puedo dar una descripción completa de hechos temporales sin usar términos perspectivistas como «ahora», «pasado» y «futuro». Siendo esto así, el requisito de completitud, aunque plausible en el caso del espacio, no lo es en el del tiempo. Por lo mismo, la idea de un hecho de perspectiva no es «idiosincrásico», sino exactamente el reflejo de lo que serían los hechos temporalmente contextualizados si la teoría A fuera verdadera.

## M. A. E. Dummett (1925-presente)

Sir Michael Dummett nació en Londres en 1925. Después de cumplir servicio militar en la Segunda Guerra Mundial, estudió en Christ Church, Oxford, antes de ser elegido como Prize Fellow de All Souls en 1950. En 1979 fue nombrado Fellow del New College y obtuvo la cátedra Wykeham de Lógica, que ocupó hasta su jubilación en 1992. Católico practicante y activista por los derechos sociales, Dummett se comprometió en campañas antirracistas en Gran Bretaña durante la década de los sesenta. Las dos mayores influencias en la filosofía de Dummett fueron Frege y Wittgenstein. Dummett escribió tres enjundiosos volúmenes sobre la filosofía del lenguaje y la filosofía de las matemáticas de Frege. También desarrolló la tesis de que los debates tradicionales sobre metafísica pueden ser aclarados, y hasta resueltos, haciendo hincapié en los debates sobre semántica. Aquí la doctrina wittgensteiniana de la naturaleza pública del significado (y la correspondiente imposibilidad de un lenguaje privado) ha sido determinante en el pensamiento de Dummett.

## ¿La teoría A o la teoría B?

Hagamos un recuento de lo alcanzado hasta ahora. Hemos descrito dos series temporales, la serie A y la serie B, y hemos puesto de manifiesto que hay dos teorías rivales del tiempo, la teoría A y la teoría B. Según la teoría A, la serie A sería fundamental para el tiempo. Según la teoría B, la serie B sería lo fundamental. McTaggart presentó dos argumentos: el primero contra la teoría B (en el que alegaba que era incompleta porque no podía dar cuenta del cambio) y el segundo contra la teoría A (en el que argüía que la serie A es contradictoria). Como McTaggart consideraba que había logrado refutar ambas teorías del tiempo, concluyó que el tiempo era irreal.

Sin embargo, hemos descubierto que ninguno de los argumentos empleados era convincente. Por un lado, hay una noción de cambio perfectamente adecuada que es definible en términos de la serie B, y, por el otro, el teórico A dispone de recursos para rebatir la inferencia de que se contradice, empleada por McTaggart. Puesto que ninguna de las dos teorías ha sido refutada, podemos plantear de nuevo la pregunta de cuál deberíamos aceptar: ¿La teoría A o la teoría B?

Las teorías A y B no son meras variantes terminológicas: ofrecen visiones de la realidad muy diferentes. Según la teoría A, la realidad está cambiando constantemente con sólo que pase el tiempo. Los eventos están continuamente intercambiando sus posiciones en la serie A. Mi nacimiento fue futuro, luego fue fugazmente presente y luego fue pasado para siempre. Esta visión dinámica de la realidad es presentada a veces en términos de un *ahora* en movimiento, que ilustra el paso del tiempo.

Hay diferentes versiones de la teoría A. Según el presentismo, sólo el presente es real. C. D. Broad piensa que el presente y el pasado son reales, pero que el futuro es irreal. Para él, la suma total de realidad se va acrecentando a medida que pasa el tiempo. McTaggart, en cambio, piensa que un teórico A debería afirmar que pasado, presente y futuro son igualmente reales[19].

No obstante, parece que sólo la primera de las dos versiones hace justicia al espíritu de la teoría A, y que la concepción de Broad es la más plausible[20]. Pertenece a la esencia de la teoría A que al presente se le asigne un puesto privilegiado, puesto que no podría ocupar si el pasado, el presente y el futuro fueran igualmente reales. El *ahora* tiene una función específica porque lleva a cabo funciones ontológicas serias: a medida que se desplaza sobre los eventos, los va invistiendo de realidad. Al sostener que el futuro es irreal, el teórico A estaría teniendo en cuenta que el futuro se halla abierto: en el presente, una serie de vías están abiertas, y el *ahora* las va clausurando todas, excepto aquella en la que efectivamente se desplaza[21].

Esta visión de la realidad contrasta radicalmente con la de la teoría B. Según la teoría B, no hay un *ahora* que se desplaza; el tiempo no fluye; y pasado, presente y futuro son igualmente reales. Aunque sabemos más sobre el pasado que sobre el futuro, ello no se debería a que el futuro sea irreal, sino a que el conocimiento es de naturaleza causal, y hay poca o ninguna causalidad retroactiva en nuestro universo. Según la teoría B, «ahora» no se refiere a una entidad dinámica, sino que es un mero término deíctico. «Ahora» simplemente se refiere al tiempo de la oración, así como «yo» se refiere a quien la pronuncia, y «aquí», al lugar de la oración. El tiempo presente no tiene privilegio alguno con respecto a otras modalidades de tiempo, como tampoco lo tengo yo con respecto a otros individuos, o el lugar que actualmente ocupo lo tiene con respecto a otros lugares.

Las teorías A y B representan, por lo tanto, concepciones de la realidad muy diferentes, incompatibles entre sí. De acuerdo con (la mejor versión de) la teoría A, el tiempo —literalmente hablando— fluye, y el futuro es irreal. De acuerdo con la teoría B, el tiempo no fluye, y pasado, presente y futuro son igualmente reales. Los teóricos A tienden a enfatizar los aspectos en que el espacio y el tiempo son asimétricos, mientras que los teóricos B consideran espacio y tiempo dimensiones análogas entre sí.

No me propongo decidir aquí cuál de las teorías A y B es la correcta. Algunos creen descubrir en las observaciones de Dummett anteriormente expuestas no sólo una defensa de la teoría A contra la objeción de McTaggart, sino un alegato positivo a favor de la teoría A y de la concepción de que el tiempo no es como el espacio. Otros asumen los desarrollos de la física moderna —y, en particular, los de la teoría especial de la relatividad— como evidencia de que no hay un *ahora* absoluto y singular. Lo que está sucediendo ahora es relativo al marco de referencia que tenga uno. Lo que es presente en un marco de referencia puede ser pasado en otro. Esto, a su vez, es entendido por algunos como si socavara la teoría A[22]. Pero la disputa sigue estando muy

viva, y demuestra cómo algunas cuestiones metafísicas pueden basarse en resultados empíricos y no sólo en reflexiones a priori.

## Observaciones finales

En este capítulo, se distinguieron dos formas de ordenar los eventos en el tiempo: se pueden ordenar en la serie A y en la serie B. Se planteó la cuestión de cuál de estas series es fundamental para el tiempo, si es que alguna lo es. La teoría A sostiene que la serie A es la fundamental, mientras que la teoría B sostiene que la serie B lo es. Examinamos dos argumentos interesantes de McTaggart contra cada teoría, y constatamos que ambos eran deficientes. La pregunta sobre cuál de ambas teorías del tiempo es la correcta sigue abierta, aunque algunas consideraciones, como veremos en el próximo capítulo, favorecen la teoría A.

## Preguntas para el estudio

- ¿Como se distingue la serie A de la serie B?
- ¿Por qué pensaba McTaggart que el cambio sólo era posible en la serie A?
- ¿En qué consiste la paradoja de McTaggart?
- ¿En qué aspectos se parece el tiempo al espacio?
- ¿Atribuye el teórico A una significación ontológica sin fundamento a verdades de perspectiva (tales como «llueve ahora»)?

### Notas

1. J. M. E. McTaggart, «Time is Not Real», reimpreso en R. C. Hoy y N. Oaklander (eds.), *Metaphysics: Classic and Contemporary Readings* (Belmont, Calif.: Wadsworth) 1991, pp. 43-44.
2. Ibíd., p. 45.
3. Ibíd., p. 45.
4. Ibíd., p. 45.
5. Citado en McTaggart, ibíd., p. 45.
6. B. Russell, *Principles of Mathematics* (Cambridge: Cambridge University Press) 2.ª edición, Allen & Unwin, 1937, sección 442.
7. McTaggart, *op. cit.*, p. 46.
8. Ibíd., p. 46.
9. Ibíd., p. 48.
10. Ibíd., p. 48.
11. Ibíd., p. 49.
12. M. Dummett, «A Defence of McTaggart's Proof of the Unreality of Time», en su obra *Truth and Other Enigmas* (Cambridge, Mass.: Harvard University Press) 1978, pp. 351-358.

13. Ibíd., p. 356.
14. Ibíd., p. 357.
15. P. Horwich, *Asymmetries in Time: Problems in the Philosophy of Science* (Cambridge, Mass.: MIT Press) 1987, capítulo 2.
16. Ibíd., p. 22.
17. Ibíd., pp. 22-23.
18. Dummett, *op. cit.*, p. 354.
19. McTaggart esboza y critica la concepción de Broad: McTaggart, *op. cit.*, pp. 49-51.
20. La concepción de Broad parece ser preferible al presentismo. Ambas concepciones concuerdan en que el futuro es irreal, pero el presentismo sostiene además que el pasado es irreal. Pero ¿qué significa decir que el pasado es irreal? No puede simplemente equivaler a negar la tesis de que los eventos pasados están sucediendo ahora, pues nadie sostiene tal cosa. Debe, entonces, consistir en la negación de la tesis de que el pasado fue real. Pero eso sería contraintuitivo, pues todos concordamos, por ejemplo, en que Sócrates bebió la cicuta, la verdad de lo cual presupone que los objetos pasados fueron reales.
21. ¿Implica cualquier teoría que niegue la realidad del futuro la falsedad del determinismo (véase el capítulo 7)? Además del presentismo, otras versiones de la teoría A (a saber, versiones que, como las de Broad o McTaggart, afirman la realidad del pasado) parecen compatibles con aseveraciones como, por ejemplo, la de que ahora sería el año 3008 y nosotros estaríamos viviendo en el pasado. Pero ¿no es absurdo que una teoría del tiempo deje abierta una posibilidad como ésa?
22. Incluso prescindiendo de la teoría de la relatividad, el carácter absoluto de la simultaneidad nunca debería haber sido considerado una verdad necesaria a priori. Michael Dummett proporciona un buen ejemplo: «Imagínese que una niebla permanente separara el Viejo Mundo del Nuevo, y que no fuera posible detectar regularidad alguna en el tiempo que se requiere para viajar de cualquiera de ellos al otro: algunas veces sólo habrían pasado unos días desde la última visita de un viajero; otras, muchos años. Entonces carecería de sentido preguntar qué había estado pasando en una mitad de la tierra mientras ocurría algún evento en la otra; y, sin embargo, cada mitad tendría su propia historia», M. Dummett, *Truth and the Past* (Nueva York: Columbia University Press) 2004, p. 86. Compárese la discusión de Dummett con la de Anthony Quinton en «Spaces and Times», reimpresa en R. Le Poidevin y M. McBeath (eds.), *The Philosophy of Time* (Oxford: Oxford University Press) 1993, pp. 203-221. Nótese que la mencionada relatividad parece compatible con la idea de que cada hemisferio tiene su propio *ahora* (es decir, que dicha relatividad no contradice frontalmente la teoría A).

## Bibliografía adicional comentada

M. Dummett, «A Defence of McTaggart's Proof of the Unreality of Time», en su obra *Truth and Other Enigmas* (Cambridge, Mass.: Harvard University Press) 1980, pp. 351-357. No se trata de una defensa de la paradoja de McTaggart, sino de un alegato a favor de que se la tome en serio. Dummett piensa que la paradoja de McTaggart, aunque no muestra que el tiempo sea irreal, nos obliga a abandonar un principio intuitivo sobre la descripción de la realidad.

P. Horwich, *Asymmetries in Time* (Cambridge, Mass.: MIT Press) 1987, capítulo 2. Una discusión clara y atractiva sobre todas las cuestiones planteados aquí. Horwich cree que la paradoja de McTaggart refuta la teoría A, pero discrepa de él sobre la posibilidad del cambio en la serie B.

R. Le Poidevin y M. MacBeath (eds.), *The Philosophy of Time* (Oxford: Oxford University Press) 1993. Contiene una útil introducción de los editores, así como reimpresiones de discusiones clásicas de problemas relacionados con el tiempo, de autores como McTaggart, Prior, Mellor, Shoemaker, Dummett, Lewis y Quinton.

J. M. E. McTaggart, *The Nature of Existence,* vol. II (Cambridge: Cambridge University Press) 1927, capítulo 33. (Reimpreso como «Time is not Real» en R. C. Hoy y L. N. Oaklander [eds.], *Metaphysics: Classic and Contemporary Readings* (Belmont, Calif.: Wadsworth) 1991, pp. 43-54.) La discusión de McTaggart ha dictado la pauta de los debates del siglo XX y XXI sobre el tiempo.

H. Mellor, *Real Time* (Cambridge: Cambridge University Press) 1981, capítulo 6 (segunda edición, 1998). Al igual que Horwich, Mellor cree que la paradoja de McTaggart socava la teoría A, y que el cambio es posible en la serie B. Una defensa vigorosa de la teoría B.

D. C. Williams, «The Myth of Passage», *The Journal of Philosophy* (1951) vol. 48, pp. 457-472. (Reimpreso en Hoy y Oaklander, *op. cit.,* pp. 55-63.) Una defensa entusiasta de la teoría B del tiempo, y del universo «bloque».

## Recursos en Internet

H. Dyke (2005) «Time, Metaphysics of», en E. Craig (ed.), *Routledge Encyclopedia of Philosophy.* Bajado el 31 de mayo de 2006 de <http://www.rep.routledge.com/article/N123>.

N. Markosian (2002) «Time», *The Stanford Encyclopedia of Philosophy (edición de invierno de 2002),* Edward N. Zalta (ed.). Bajado el 31 de mayo de 2006 de <http://plato.stanford.edu/archives/win2002/entries/time>.

R. Le Poidevin (2005) «Presentism», en E. Craig (ed.), *Routledge Encyclopedia of Philosophy.* Bajado el 31 de mayo de 2006 de <http://www.rep.routledge.com/article/N120>.

L. Sklar (1998) «Time», en E. Craig (ed.), *Routledge Encyclopedia of Philosophy.* Bajado el 31 de mayo de 2006 de <http://www.rep.routledge.com/article/Q107>.

# 6. El tiempo: tres problemas

## Introducción

Me propongo abordar en el presente capítulo tres problemas que se refieren al tiempo, dos de los cuales tienen implicaciones para el debate sobre la naturaleza del tiempo. Discutiré el problema de Prior conocido como «¡Menos mal que eso se acabó!» [*Thanks goodness that's over*], el de Shoemaker sobre la posibilidad del tiempo sin el cambio y el de Lewis sobre la posibilidad de viajar en el tiempo. Tanto el argumento de Prior como la posibilidad del tiempo sin el cambio proporcionan apoyo a la teoría A.

## El problema de Prior

En 1959 Prior planteó un problema con respecto a la teoría B del tiempo, que expresó en los términos siguientes:

> Uno dice, por ejemplo, «menos mal que eso acabó», y no sólo sucede que esta frase, cuando se profiere, resulta perfectamente clara sin necesidad de especificar la fecha de su emisión, sino que dice algo imposible de expresar usando sólo una cópula temporalmente descontextualizada y una fecha. Ciertamente, no significa lo mismo que «menos mal que la fecha en que eso acabó es el viernes 15 de junio de 1954», incluso si la frase ha sido pronunciada en esa fecha. (Ni, mucho menos, significa lo mismo que «menos mal que la conclusión de eso es contemporánea con esta emisión». Pues, ¿por qué habría alguien de sentirse aliviado por ese hecho?)[1]

## El viaje en el tiempo

Muchas series de televisión (por ejemplo, algunos episodios de *Star Trek*), películas (por ejemplo, *Terminator*), así como novelas de ciencia ficción (por ejemplo, *La máquina del tiempo,* de H. G. Wells), nos presentan a personajes que viajan en el tiempo. Lo que distingue a un viajero del tiempo de una persona ordinaria es que el tiempo personal de un viajero semejante difiere del tiempo externo. Un viajero del tiempo puede viajar mil años al pasado o al futuro en sólo una hora de su tiempo personal. Esto es, su cabello crece en la proporción en que lo hace normalmente en una hora, su reloj de pulsera le indica que ha transcurrido sólo una hora, etc. Y, sin embargo, han pasado realmente mil años. Un mundo en el que hay viajes en el tiempo es, cuando menos, extraño. Si el viaje es hacia el pasado, involucra la causalidad retroactiva, y puede contener anomalías metafísicas, tales como los circuitos causales. Algunos piensan que si el viaje hacia el pasado fuera posible, un viajero en el tiempo que quisiera cambiar el pasado podría retroceder temporalmente y modificarlo. Ésta es la premisa de la película *Terminator*. Pero se trata de una premisa falsa. Un propósito como ése es irrealizable, pues «cambiar el pasado» es una descripción contradictoria. Así pues, hay una incongruencia en el corazón mismo de este filme.

Es decir, la teoría B no dispone de los recursos necesarios para dar a oraciones como «menos mal que eso acabó» el contenido deseado. Tales oraciones pueden tener dicho contenido únicamente si la teoría A es verdadera.

Sin embargo, Prior se ha equivocado ligeramente al formular el problema. La teoría B es una teoría sobre la naturaleza del tiempo, no una teoría sobre el significado de oraciones temporalmente contextualizadas [*tensed sentences*] (es decir, oraciones que contienen términos de la serie A, como «pasado», «presente» o «futuro»). La teoría B no pretende que una oración temporalmente contextualizada, por ejemplo mi oración de «la muerte de Hitler sucedió en el pasado» signifique lo mismo que su contrapartida temporalmente descontextualizada «esa oración de "la muerte de Hitler sucedió en el pasado" ocurrió después de la muerte de Hitler». La tesis de la teoría B es que las verdades temporalmente contextualizadas (aquellas que se valen de «pasado», «presente», «futuro», así como de otros términos de la serie A) son **reducibles** a verdades temporalmente descontextualizadas (aquellas que se expresan usando únicamente términos y relaciones de la serie B). De forma análoga, también los hechos temporalmente contextualizados serían reducibles, según la teoría B, a hechos temporalmente descontextualizados. En general, la reducción de un hecho a otro no necesita involucrar una pretensión de sinonimia entre oraciones.

Afortunadamente, el problema de Prior, como ha observado Hugh Mellor, puede ser reformulado de manera tal que involucre directamente a la teoría B:

Supongamos que tú acabas de tener una experiencia dolorosa, como una jaqueca. Ahora que ésta ha terminado, dices con alivio «menos mal que eso acabó». ¿Por qué te sientes aliviado? Por el hecho de que la jaqueca ya no siga siendo una experiencia presente, es decir, por el hecho de que ahora pertenezca al pasado. Ésa es presumiblemente la razón de que hicieras tu observación una vez que el dolor pasó, y no, en cambio, antes de que ocurriera, ni mientras ocurría. ¿Puede esto [...] seguir siendo explicado [dentro de la teoría B]?[2]

He aquí, pues, el meollo del problema. En la teoría B, todos los hechos temporales están temporalmente descontextualizados (podremos seguir hablando, si lo deseamos, de hechos temporalmente contextualizados, pero éstos serán reducibles a aquéllos). Los hechos temporalmente descontextualizados son fijos e invariables. Valen todo tiempo. Exclamar «¡menos mal!» no podría, por lo tanto, ser más apropiado respecto de un hecho temporalmente descontextualizado *que respecto de otro*. Y, sin embargo, es apropiado usar esa expresión cuando una jaqueca ha pasado, pero no cuando ésta se halla presente o va a producirse en el futuro. Eso significa que, al exclamar «¡menos mal!» cuando la jaqueca ya pasó, estoy expresando alivio por un hecho temporalmente contextualizado (el hecho de que la jaqueca haya pasado); no estoy, esto es, exclamando «¡menos mal!» por un hecho temporalmente descontextualizado; de ahí que los hechos temporalmente contextualizados no sean reducibles a hechos temporalmente descontextualizados, y que la teoría B sea falsa.

La respuesta que da el propio Mellor a esta versión del problema de Prior consiste en afirmar que cuando digo «¡menos mal que eso acabó!», después de haber sufrido un fuerte dolor de cabeza, no estoy expresando alivio por un hecho temporalmente contextualizado como el de que mi jaqueca pertenezca al pasado, sino simplemente expresando alivio (y no sobre algo en particular). Sin embargo, esta respuesta es poco convincente. Incluso si el alivio no siempre tiene un objeto intencional (hay, en efecto, casos en que uno siente alivio, pero no sobre algo en particular), lo usual es que el alivio tenga un objeto intencional, como sin duda ocurre en el caso que estamos considerando. Cuando exclamo «¡menos mal que eso acabó!» después de sufrir una jaqueca, es porque me siento aliviado sobre algo en particular, a saber, por el hecho de que mi jaqueca haya pasado. Por ende, la redescripción que hace Mellor de este caso es poco plausible.

Lo que el ejemplo de Prior demuestra es que algunas de nuestras disposiciones presuponen hechos temporales cambiantes. Siento alivio, porque el hecho de que mi jaqueca estaba en el presente ha dado paso al hecho de que esa misma jaqueca está ahora en el pasado. Puesto que en la teoría B todos los hechos temporales son temporalmente descontextualizados e invariables, el teórico B no podría dar cuenta de una disposición como ésta.

Es más, el punto suscitado por Prior es generalizable. El problema no es sólo que la teoría B no pueda dar cuenta de expresiones como «¡menos mal que eso acabó!». Tampoco puede dar cuenta de muchas de las disposiciones que tenemos hacia nuestras propias experiencias pasadas y futuras. Por ejemplo, sentimos un creciente alivio cuanto más se aleja en el pasado algún

episodio bochornoso. ¿Cómo podríamos explicar eso en términos de la serie B? ¿Por qué habría de sentirme reconfortado por el hecho de que, por ejemplo, 2008 esté situado cinco años después de los eventos bochornosos de 2003? ¿O de que 2009 esté situado seis años después de dichos eventos? Trivialidades como éstas mal podrían servir de asidero a nuestros sentimientos de alivio.

Por otra parte, considérese el hecho de que solemos preocuparnos más por los dolores futuros que por los pasados. Preferimos con mucho el haber dejado atrás una experiencia dolorosa que aún tener que encararla. Estamos, como dice Parfit, «prejuiciados hacia el futuro»[3]. ¿Cómo explicar, desde la teoría B, esta inclinación? Las relaciones «anterior a» y «posterior a» son perfectamente simétricas. ¿Por qué el hecho de que un dolor sea posterior al momento de la oración, y no anterior a ella, debería ser una razón para que nos importe más?

El teórico A podría, al menos, decir lo siguiente. El tiempo tiene una dirección intrínseca que va desde el pasado hacia el futuro, y esto es lo que subyace a la asimetría típica de nuestra manera de preocuparnos. Es razonable interesarse más por los dolores futuros, porque éstos nos salen al paso; y es razonable interesarse menos por los dolores pasados, porque éstos se alejan más y más de nosotros. Pero ¿qué podría decir el teórico B? El problema de Prior es sólo uno entre aquellos que ocasionan dificultades especiales en la teoría B.

### A. N. Prior (1914-1969)

Arthur Norman Prior nació en Nueva Zelanda. Fue catedrático de Filosofía en la Universidad de Manchester y más tarde Fellow del Balliol College de Oxford. Murió repentinamente en 1969, sólo tres años después de llegar a Balliol. Prior hizo importantes contribuciones al desarrollo de la lógica temporal (*tense logic*) y a la metafísica del tiempo. Exploró las analogías entre la lógica de lo necesario y lo posible, por un lado, y la lógica del pasado, presente y futuro, por otro. Consideró que los operadores temporales «fue el caso que p» y «será el caso que p» eran análogos a los operadores modales «es necesario que p» y «es posible que p». Así, creyó que la manera más perspicaz de representar una oración ordinaria conjugada en pretérito, como «Guillermo fue calvo», era «fue el caso que: Guillermo es calvo». Prior pensaba que las proposiciones temporalmente contextualizadas eran necesarias para la expresión de hechos temporales, y, por consiguiente, se le podría considerar un teórico A del tiempo. Prior recurría frecuentemente a las obras de los lógicos antiguos y medievales, y fue, desde todo punto de vista, un maestro incansable y estimulante.

## Tiempo sin cambio

¿Qué conexión existe entre el tiempo y el cambio? Es indudable que el cambio implica tiempo: el cambio es un proceso temporal. Pero ¿implica el tiem-

po el cambio? Es decir, ¿es posible que haya un periodo de tiempo en el que nada cambie en el universo entero?[4] Nótese que no estamos preguntando si ha habido o habrá en nuestro universo un periodo de tiempo carente de cambio. Simplemente estamos preguntando si la idea en sí misma tiene sentido, si es, lógica o conceptualmente, posible que exista un periodo de tiempo carente de cambio. Y, si existiera, ¿podría la gente de un mundo como ése predecir cuándo ha transcurrido un periodo de tiempo carente de cambio? En su interesante artículo «Tiempo sin cambio», Sydney Shoemaker elabora un ingenioso argumento para responder afirmativamente a esta pregunta[5].

Algunos filósofos famosos han insistido en que no puede haber tiempo sin cambio: Aristóteles, Hume y McTaggart son sólo tres ejemplos ilustres. Sin embargo, como se recordará del capítulo anterior, McTaggart contaba como cambio el que un evento se alejara en el pasado. Según esa concepción, el mero paso del tiempo contaría como un cambio, y, por consiguiente, el tiempo sin cambio sería imposible por definición.

En su discusión, Shoemaker no considera a estos **cambios McTaggart** como cambios genuinos, y entiende «cambio» como «cambio ordinario»: un objeto cambia cuando se modifica respecto de su color, tamaño, peso, velocidad, posición en el espacio, etc. La pregunta de si puede haber tiempo sin cambio, así entendida, deja de tener una respuesta automática. Esa pregunta hace referencia, además, al sentido en que utilizaron la palabra «cambio» Aristóteles y Hume cuando negaron que pudiera haber un periodo de tiempo sin cambio.

Pero ¿por qué pensar que el tiempo implica necesariamente la presencia del cambio (en el sentido ordinario de «cambio»)? Shoemaker cita la razón aducida por Aristóteles:

> [...] cuando el estado de nuestras propias mentes no cambia en absoluto, o no hemos notado que cambia, no nos damos cuenta de que el tiempo ha transcurrido, así como no se daban cuenta de ello, tras despertar, aquellos héroes que, según la leyenda, dormían en Sardinia; pues éstos conectaban el «ahora» anterior con el posterior, y los unían, y, no advirtiendo el intervalo, dejaban a éste por fuera[6].

Aunque el sentido de estas observaciones no sea completamente claro, ellas parecen expresar la siguiente concatenación de ideas. No puede haber tiempo sin cambio porque nuestra consciencia del paso del tiempo involucra necesariamente la consciencia del cambio (que se da en nuestras propias mentes o en el mundo observable).

No obstante, este argumento no es convincente. Para empezar, adolece de un fallo estructural. Su conclusión es metafísica (el tiempo implica necesariamente el cambio), pero su premisa es epistémica (concierne a nuestra consciencia del paso del tiempo). ¿Cómo podría una conclusión metafísica seguirse de una premisa epistémica? ¿Está presuponiendo Aristóteles que para que el tiempo transcurra tenemos que estar conscientes de ello? Además, la premisa de Aristóteles —el que nuestra consciencia del paso del tiempo involu-

cre necesariamente nuestra consciencia del cambio— es ambigua. Ignora la distinción entre consciencia directa e indirecta. Es obvio que yo no puedo estar directamente consciente del tiempo si no hay cambio, ya que la propia continuidad de mi consciencia constituye un cambio; pero ¿por qué no podría tener yo evidencia indirecta de que ha transcurrido un periodo determinado de tiempo carente de cambio? Para usar la analogía de Shoemaker, no puedo verificar directamente «en un tiempo t, el universo no contiene mente alguna en t», ya que todo intento de hacerlo se refutaría a sí mismo; pero puedo ciertamente tener evidencia indirecta, a partir de la cosmología, que justifique mi aceptación de dicho enunciado[7].

Shoemaker proporciona, a renglón seguido, un ejemplo de un mundo en el que hay un periodo de tiempo sin cambio y en el que la gente tiene una buena razón para creer que ha transcurrido cierto periodo de tiempo sin cambio. Se imagina un mundo dividido en tres regiones espaciales: A, B y C. Hay una interacción típica entre la gente de esas tres regiones, y la gente se puede mover libremente entre ellas. Pero se produce la siguiente anomalía: cada cierto tiempo una de las regiones «se paralizaría» durante un año. Además, cuando A se paraliza, la gente en B y C puede observar que no ocurren eventos en A. Pasado el año, todo en A vuelve a recuperar su movimiento. La gente en A sigue conversando entre sí como si no hubiera habido interrupción. Por supuesto, las cosas le parecerán extrañas a cualquier ocupante de A que haya estado observando las cosas que pasaban en B o C antes de la paralización. Una vez concluida ésta, tendrá la impresión de que muchas cosas ocurren instantáneamente. Sin embargo, cuando los ocupantes de B y C le expliquen lo que en realidad ha sucedido, la extrañeza del ocupante de A disminuirá un poco. Entenderá que lo que sucedió en la región A también sucede periódicamente en las otras regiones, y que tales paralizaciones pueden ser verificadas cada vez por los ocupantes de las regiones no paralizadas.

Hasta ahora no tenemos un caso de tiempo sin cambio. Del modo en que hemos venido describiendo nuestro mundo imaginario, siempre que una región está paralizada, hay otras regiones no paralizadas en las que ocurren cambios. Pero supongamos que los miembros más perspicaces del mundo en cuestión comienzan a notar una cierta regularidad en las paralizaciones: A se paraliza cada tres años, B, cada cuatro años, y C, cada cinco años. A partir de esta información, esas personas pueden inferir, por simple cálculo aritmético, que habrá una paralización global cada sesenta años, en la que A, B y C se paralicen simultáneamente durante todo un año[8]. Ahora tenemos lo que buscábamos: un mundo en el que hay un periodo de tiempo sin cambio alguno, y cuyos habitantes son, sin embargo, capaces de predecir este último.

El mundo de Shoemaker parece posible, pero plantea una pregunta insoslayable: ¿Qué sería lo que ocasionase el final de una paralización? En el caso de una paralización localizada, como, digamos, la de A, podrían ser eventos en B o C los que ocasionasen el cese de la paralización de A. Pero, ¿qué ocasionaría el final de la paralización global? Si el universo entero estuviera paralizado,

¿qué podría hacer que dejara de estarlo? No, por cierto, eventos que sucedieran durante la paralización global, pues no existen tales eventos. Pero la causa tampoco podría ser un evento previo a la paralización, ni simultáneo con su inicio, pues, de lo contrario, la paralización habría terminado nada más comenzar.

Por muchas vueltas que le demos al asunto, parece que tenemos que renunciar al principio de **«no hay acción a distancia temporal»**. Esto es, la clase de causalidad vigente en nuestro mundo imaginario tendría que violar el principio siguiente:

(P) Si un evento está causado, entonces cualquier intervalo temporal que lo preceda, por muy pequeño que sea, contendrá una causa suficiente de su ocurrencia.

Como especifica Shoemaker: «Suponer que (P) es falsa equivale a suponer que un evento puede ser causado directamente, es decir, no a través de una cadena causal, por un evento que ocurrió un año antes; o que un evento puede ser causado por el hecho de que tal y tal cosa haya sido el caso durante el periodo de un año»[9]. Aunque solemos presuponer que nuestro mundo está gobernado por el principio (P), Shoemaker no ve obstáculo conceptual alguno a este último. Cree que podría haber un mundo como el que hemos imaginado, que contuviera una relación identificable como «causalidad», la cual fuera violatoria del principio (P).

En un mundo en el que (P) se infringiese, el mero paso del tiempo tendría eficacia causal. Dado que los cambios causalmente eficaces son genuinos, los cambios meramente temporales de dicho mundo serían genuinos. En otras palabras, los cambios McTaggart, que antes Shoemaker había excluido del ámbito de los cambios reales, tendrían que ser considerados genuinos, al menos en los mundos infractores de (P). Pues el paso del tiempo en dicho mundo sería causalmente eficaz. Pero entonces, el mundo imaginario de Shoemaker no sería un mundo tal que albergase un tiempo sin cambio.

A pesar de ello, Shoemaker tiene éxito en describir un mundo dotado de un tiempo sin cambio (ordinario) alguno, en el que los periodos de tiempo carente de cambio pueden ser predichos por sus habitantes. Lástima que en ese mundo el cambio ordinario no sea la única clase de cambio genuino. Ese mundo admitiría periodos de tiempo sin cambio ordinario, pero no periodos de tiempo sin cambio alguno.

Para concluir, diremos que la tesis tan plausiblemente defendida por Shoemaker —respecto de que puede haber periodos de tiempo sin cambio ordinario— plantea una dificultad para el teórico B. ¿Cómo es posible para el teórico B que pase un año de tiempo carente de cambio, si no hay eventos durante ese año que puedan hallarse en relaciones de la serie B con otros eventos? Recuérdese que el paso del tiempo consistía para el teórico B sólo en eventos ordenados en la serie B. Al igual que el problema de Prior, esta consideración habla en contra de la teoría B y a favor de la teoría A.

### Tiempo sin cambio

Numerosos filósofos (Aristóteles, Hume y McTaggart, por sólo mencionar a tres) creyeron que el tiempo implica cambio. Es decir, que para que el tiempo transcurra, alguna porción de cambio, por pequeña que sea, tiene que darse en algún rincón del universo. Dada la forma en que McTaggart entendió «tiempo» y «cambio», es una verdad definicional, y no un argumento, el que el tiempo implique cambio. Aristóteles, por su parte, sí ofreció un argumento, aunque no muy bueno, para probar la imposibilidad de que haya tiempo sin cambio. Adujo que no puede haber tiempo sin cambio porque la consciencia del paso del tiempo implica necesariamente la consciencia del cambio. Pero este argumento pasa por alto la posibilidad de que el tiempo pase sin que nos demos cuenta de ello y no haya cambio alguno durante ese periodo. Por supuesto, no podemos estar directamente conscientes de un tiempo desprovisto de cambio, ya que una consciencia sostenida es ella misma un cambio. Pero podríamos tener evidencia indirecta de que ha transcurrido un periodo de tiempo carente de cambio. El mundo imaginario de Shoemaker ilustra muy bien esa posibilidad. Al mostrar cómo la gente de su mundo podría llegar a creer razonablemente que un año de tiempo ha transcurrido sin que suceda cambio alguno, Shoemaker provee evidencia para la tesis de que el tiempo sin cambio es una posibilidad lógica.

## Viaje en el tiempo

Para muchos, el tema del viaje en el tiempo es el más excitante de la filosofía del tiempo. Desde un punto de vista psicológico, las preguntas más fundamentales al respecto son: ¿Es posible viajar en el tiempo?, y ¿cómo sería un mundo en el que se pudiera viajar en el tiempo? Al formular estos interrogantes, no estamos presuponiendo que en nuestro mundo haya ocurrido, o vaya a ocurrir, un viaje en el tiempo. Sencillamente estamos preguntando si tal viaje es lógicamente posible. Y por «posible» nos referimos a posible en el sentido más amplio del término: esto es, lógica o conceptualmente posible. No nos interesa aquí si el viaje en el tiempo es o no físicamente posible (es decir, compatible con las actuales leyes de la naturaleza), aunque ésta sea una pregunta interesante; nos interesa sólo si un viaje en el tiempo tiene sentido. Todos disfrutamos de las novelas y de las películas que incluyen viajeros en el tiempo; pero ¿corresponden esas historias a posibilidades reales? Para responder a esta pregunta, me apoyaré en el excelente artículo de David Lewis titulado «Las paradojas del viaje en el tiempo»[10].

En primer lugar, debemos definir «viaje en el tiempo». Como Lewis dice: «inevitablemente, [el viaje en el tiempo] implica una discrepancia entre un tiempo y otro»[11]. El viaje en el tiempo puede darse hacia el pasado o hacia el futuro: en ambos casos, puede pasar que la travesía del viajero del tiempo

dure, digamos, una hora, y que, sin embargo, éste retroceda cientos de años hacia el pasado o avance cientos de años hacia el futuro. La idea de una discrepancia entre un tiempo y otro puede sonar absurda; sin embargo, Lewis logra ser coherente

[...] distinguiendo el tiempo mismo, *el tiempo externo* (como lo llamaré), del tiempo personal de un viajero del tiempo (aproximadamente: lo medido por su reloj de pulsera). Supongamos que su travesía dura una hora de tiempo personal: su reloj marcará una hora más cuando llegue a su destino que la que marcaba cuando inició su viaje. Pero tarda en llegar más de una hora[12].

Es importante observar que la distinción entre el tiempo personal y el externo no equivale a dos dimensiones del tiempo. Lewis dice del tiempo personal que

[...] realmente no es un tiempo [específico], ya que desempeña en la vida (del viajero del tiempo) la misma función que el tiempo desempeña en la vida de una persona común. [...] Podríamos comparar los intervalos de tiempo externo con las distancias que recorre un cuervo en su vuelo; y los intervalos de tiempo personal, con las distancias a lo largo de un camino serpenteante. La vida del viajero en el tiempo es como los rieles de un tren que atraviesa una zona montañosa [...] no estamos aquí frente a dos dimensiones independientes. Así como la distancia a lo largo de la vía del tren no constituye una cuarta dimensión espacial, así el tiempo personal de un viajero en el tiempo no es una segunda dimensión en el tiempo[13].

La distinción entre el tiempo personal y el tiempo externo nos permite dar sentido al discurso sobre la identidad personal en el caso de un viajero del tiempo. Queremos poder decir que el viajero en el tiempo que entra en la máquina del tiempo en el año 2009 es *la misma persona* que el hombre que sale de la máquina, pasada una hora de su tiempo personal, en el año 1900. A menudo se cree que las relaciones de **continuidad mental y/o corporal** son lo que explica la identidad a través del tiempo. (Véase el capítulo 8.) En el caso de alguien que no viaja en el tiempo, la frase «a través del tiempo» no es equívoca. Pero en el caso de un viajero del tiempo, puede querer decir tanto «a través del tiempo externo» como «a través del tiempo personal». Lewis sugiere que deberíamos considerar la identidad personal de un viajero en el tiempo como si estuviera constituida por una continuidad mental y/o corporal con respecto al tiempo personal. Esto nos permitiría hacer justicia a la noción intuitiva de que la persona que entre en la máquina del tiempo en 2009 es la misma persona que sale de la máquina en 1900[14].

## David Lewis (1941-2002)

David Lewis dio clases por corto tiempo en la Universidad de Los Ángeles, California, antes de trasladarse a Princeton en 1970. Escribió sobre muchas áreas de la filosofía, pero se le conoce especialmente por sus trabajos sobre los condicionales contrafácticos y la filosofía de la modalidad (la posibilidad y la necesidad). De acuerdo con Lewis, un condicional contrafáctico de la forma «si A hubiera sido el caso, B habría sido el caso» es verdadero si algún mundo posible en el que tanto A como B son verdaderos está «más próximo» a nuestro mundo que cualquier otro mundo en el que A y ~ B sean verdaderos. Lewis asume la «proximidad» como una cuestión de semejanza, y adopta una visión realista de los mundos posibles: los mundos posibles existen exactamente del mismo modo en que lo hace nuestro mundo. No hay nada especial o privilegiado acerca del mundo actual, ya que cada mundo es actual para sus habitantes. Algunos han objetado que la «proximidad» no puede ser entendida en términos de semejanza; otros, que la concepción realista de Lewis sobre objetos posibles y mundos posibles es inverosímil (ya que nos invita a que creamos en montañas de oro y asnos parlantes que existen exactamente de la misma manera en que lo hacen nuestras montañas y nuestros asnos).

Hemos trazado, pues, una importante distinción entre el tiempo personal y el tiempo externo, que nos permite, al menos *prima facie,* dar cierto sentido al viaje en el tiempo. Sin embargo, podría haber perplejidades y paradojas involucradas en la idea de un viaje en el tiempo y en la de agentes (personas) que viajan en el tiempo. Estas perplejidades y paradojas tienden a ser presentadas con respecto a viajes hacia el pasado, y no con respecto a viajes hacia el futuro. La pregunta que se nos plantea es la de si estas perplejidades y paradojas representan objeciones genuinas a la posibilidad de un viaje en el tiempo. Argumentaré que no lo hacen, y que, por lo tanto, no existe barrera conceptual alguna con respecto a un viaje hacia el pasado o hacia el futuro. Expondremos a continuación algunas dificultades.

### Causalidad retroactiva

Viajar en el tiempo hacia el pasado involucra la causalidad retroactiva [*backwards causation*] (con respecto al tiempo externo). Como dice Lewis refiriéndose a alguien a punto de viajar hacia el pasado: «puedes darle un golpe en la cara antes del viaje, y hacer que su ojo se hinche hace siglos»[15]. Asimismo, un viajero del tiempo que enciende su máquina del tiempo en el año 2009 causa que la máquina del tiempo llegue a su destino en 1900. ¿Constituye una objeción a la posibilidad misma de viajar hacia el pasado el que esto presuponga la causalidad retroactiva? Sólo si es incoherente la idea de causalidad retroactiva, es decir, la idea de un efecto que precede a su causa. Prescindiendo de teo-

rías que estipulan simplemente que un evento vale como causa sólo si precede a su efecto, ninguna de las teorías principales de la causalidad excluye la posibilidad de la causalidad retroactiva. Por lo tanto, no es razonable descartar un viaje hacia el pasado por el mero hecho de que éste haga inevitable la causalidad retroactiva. Es más, a falta de otras objeciones que excluyan los viajes hacia el pasado, podemos incluso decir que la posibilidad del viaje hacia el pasado es un argumento a favor de la posibilidad de causalidad retroactiva.

## Circuitos causales

Aparte de la causalidad retroactiva, el viaje hacia el pasado también da lugar a la posibilidad de circuitos causales [*causal loops*]. Los circuitos causales son...

> [...] cadenas causales cerradas en las que algunas de las secuencias causales tienen una dirección normal, y otras, la dirección contraria [...] Cada evento del circuito posee una explicación causal, y está causado por eventos situados en otros lugares del circuito. Ello no significa que el circuito como un todo esté causado o sea explicable. Podría no ser así[16].

Se puede construir un buen ejemplo de circuito causal en términos de transferencia de información. Imagínese a un viajero del tiempo que retrocede unos cuantos años y habla con su yo anterior. Hablan del viaje en el tiempo y

> [...] en el transcurso de la conversación, el yo más viejo le dice a su yo más joven cómo construir una máquina del tiempo. Se trata de una información a la que no es posible acceder de otro modo. Su yo más viejo sabía cómo hacer la máquina porque a su yo más joven le había sido dada la información y éste la había retenido [en su memoria]. Y el yo más joven se enteró de la información por haber conversado con su yo más viejo [que se la dio]. Pero ¿cuál fue el verdadero origen de la información? ¿Por qué ocurrió todo el asunto? Sencillamente, no hay respuesta[17].

¿Son imposibles los circuitos causales? Si así fuera, entonces viajar hacia el pasado debería serlo también. Sin embargo, no hay razones para pensar que los circuitos causales sean imposibles. No vacilamos en concebir la posibilidad de que haya eventos inexplicables o no causados: «Dios, el Big Bang, todo el pasado infinito del universo, o la decadencia de un átomo de tritio»[18]. Si estas cosas son posibles, ¿por qué no los circuitos causales? La posibilidad de circuitos causales muestra que los mundos en los que la gente viaja hacia el pasado son extraños y muy diferentes del nuestro, pero no que son imposibles.

## La paradoja del abuelo: ¿Puede un viajero del tiempo cambiar el pasado?

Una de las objeciones más conocidas que se formulan contra el viaje en el tiempo es la posibilidad de que un viajero del tiempo cambiase el pasado. Como es imposible cambiar el pasado, se concluye que también viajar hacia el pasado es imposible.

No cabe duda de que cambiar el pasado es imposible. Cambiar el pasado es hacer que sea verdadero que sucedió algún evento que no sucedió, o hacer que sea verdadero que no sucedió algún evento que sucedió. Pero nunca puede ser verdadero el que algún evento sucedió y no sucedió: ni siquiera Dios podría hacer que eso fuera verdadero. No hay nada especial sobre el pasado a este respecto. Es igualmente imposible cambiar el presente o el futuro. Nadie puede hacer que sea el caso que un evento suceda *y* no suceda, o que algún evento vaya a suceder *y* no vaya a suceder. Por supuesto, podemos modificar o crear el futuro (decidiendo actuar de una manera u otra), pero no podemos cambiarlo en el sentido recién definido. Es indudable que aquellos que creen en la posibilidad de viajar hacia el pasado están obligados a aceptar la posibilidad de modificar o crear el pasado. Pero ¿tienen también que aceptar la posibilidad de que un viajero en el tiempo modifique el pasado?

Lewis cree que no, y, mediante el ejemplo siguiente, comienza por delinear la tesis de que un viajero del tiempo es capaz de cambiar el pasado:

> Considérese el caso de Tim. Detesta a su abuelo, cuyo éxito en la fabricación de municiones fue el origen de la fortuna familiar que costeó la máquina del tiempo que Tim construyó. No hay nada que éste desee más que matar a su abuelo, pero desgraciadamente llegó muy tarde. El abuelo murió en su cama en 1957, cuando Tim era sólo un niño. Pero cuando Tim construye luego su máquina del tiempo y viaja al año 1920, se da cuenta de que, después de todo, no es demasiado tarde. Compra un rifle y pasa largas horas practicando el tiro al blanco; sigue a su abuelo como si fuera su sombra, para descubrir su itinerario cotidiano[19].

Tim puede matar a su abuelo: tiene un rifle de larga distancia; también tiene buena puntería, las condiciones del clima son ideales, etc. Pero Tim no puede matar a su abuelo: el abuelo murió en su cama en 1957, por lo que no puede haber muerto en 1920. La consistencia exige que Tim, a pesar de sus esfuerzos, no logre matar a su abuelo. ¿Por qué falla? «Por una razón muy sencilla. Quizás algún ruido lo distraiga en el último momento; quizás yerre el blanco, a pesar de su entrenamiento; quizás le fallen los nervios; quizás incluso lo asalte un ataque repentino de misericordia.»[20] Por lo tanto, concluye Lewis, es equivocado pensar que un viajero del tiempo pueda cambiar el pasado.

Aquí cabe hacer tres observaciones. En primer lugar, parece que Lewis no ha hecho otra cosa que reemplazar una contradicción por otra. ¿Acaso no se ha sustituido sencillamente la contradicción «el abuelo de Tim muere en 1920 y en 1957» por la contradicción «Tim puede y no puede matar a su abuelo»?

En ese caso, no sería extraño que el viaje hacia el pasado siguiera implicando una contradicción.

Lewis tiene una buena réplica a esta objeción. No hay contradicción alguna, dice, ya que «puede» es equívoco. En relación con una serie de hechos (relativos, por ejemplo, al rifle de Tim, a su puntería, a las condiciones climáticas, etc.), Tim puede matar a su abuelo. Pero en relación con otros (incluyendo el hecho de que el abuelo no murió en 1920), no puede hacerlo. Sólo habría una contradicción si Tim pudiera y no pudiera matar a su abuelo en relación con la misma serie de hechos. Pero eso no es algo que tenga que aceptar quien defiende la posibilidad de viajar en el tiempo.

En segundo lugar, la historia de Lewis se refuta causalmente a sí misma: al fin y al cabo, Tim está tratando de eliminar una de las causas de su propia existencia. Esto hace que el proyecto de Tim sea doblemente imposible. No sólo se refuta a sí mismo, sino que es un intento de cambiar el pasado. Pero el aspecto de la autorrefutación no es esencial a la solución de Lewis. Aunque lo que se propusiera hacer Tim fuera asesinar al socio comercial de su abuelo, también fallaría, y por el mismo tipo de razón (asumiendo, por ejemplo, que el socio vivió hasta 1950).

En tercer lugar, Lewis tiene razón en que la historia de Tim, para ser consistente, tiene que ajustarse a las pautas sugeridas por él. Tim falla de algún modo: su rifle se atasca, mata a otra persona, yerra el blanco, etc. Pero esta solución nos deja insatisfechos. Supóngase que, día tras día, Tim trata de asesinar a su abuelo, y que algo siempre sale mal. Un día hace demasiado viento, al siguiente el dedo se le resbala cuando va a apretar el gatillo, otro día su arma se atasca, y así sucesivamente. Cada percance podrá tener su explicación, pero no la secuencia completa de ellos. Dado que Tim posee, según los criterios usuales, el deseo y los medios necesarios para matar a su abuelo, la secuencia total de percances constituiría una sucesión de eventos altamente improbable. Lo mismo valdría para cualquier otro viajero del tiempo empeñado en «cambiar el pasado». Pero aunque eso resultase extraño, quizás no sería imposible. Secuencias improbables como éstas no demuestran que el mundo de un viajero en el tiempo sea imposible, aunque sí confirman lo que ya sabemos, a saber, que un mundo semejante sería muy diferente del nuestro.

Finalmente, vale la pena destacar que el tema del viaje en el tiempo se intersecta con el debate entre la teoría A y la teoría B que expusimos en el capítulo anterior. Si el presentismo es cierto, y sólo el presente es real, viajar en el tiempo debería ser imposible. Si tanto el pasado como el futuro son irreales, ¿adónde podría ir el viajero del tiempo? Pero si la concepción de Broad es la correcta —respecto de que el pasado y el presente son reales, mientras que el futuro es irreal—, entonces viajar al futuro debería ser imposible. (Por lo tanto, si recibiéramos la visita de viajeros provenientes del futuro, ello refutaría estas dos versiones de la teoría A.) Sólo la teoría B, que sostiene que pasado, presente y futuro son igualmente reales, admite la posibilidad de viajar hacia el pasado y hacia el futuro. Puede que algunos entiendan los resultados de nuestra dis-

cusión sobre viajar en el tiempo como evidencia a favor de la teoría B, pero con ello quizás no estén sino presuponiendo lo que deberían demostrar. Es preferible concluir que si el pasado es real, entonces no hay barrera conceptual alguna para viajar hacia al pasado (y lo mismo vale para el futuro).

## Observaciones finales

Tras haber examinado tres problemas interesantes sobre el tiempo (el problema de Prior, el problema del tiempo sin cambio y el problema del viaje en el tiempo), hemos llegado a algunas conclusiones interesantes. El análisis del problema de Prior permitió inferir un argumento ingenioso a favor de la teoría A. Por otro lado, vimos cómo Shoemaker describía exitosamente un mundo posible cuyos habitantes pueden tener constancia de que durante un año ha transcurrido un tiempo carente de cambio (ordinario). Este resultado favorece, además, la teoría A, ya que no resulta obvio cómo la teoría B podría explicar que el tiempo transcurra a pesar de que no haya evento alguno que pueda ser ordenado en la serie B. Finalmente, nuestra discusión del viaje en el tiempo mostró que, si aceptamos que el pasado es real, un mundo en el que ocurran los viajes hacia el pasado será muy diferente del nuestro, pero, aun así, posible.

## Preguntas para el estudio

- ¿Puede el teórico B argumentar razonablemente que nuestras actitudes temporalmente predispuestas son irracionales?
- ¿Puede el teórico A proporcionarnos una explicación satisfactoria de por qué nos importan más nuestras experiencias futuras que nuestras experiencias pasadas?
- ¿Es posible generar un correlato espacial del razonamiento de Prior (como «menos mal que eso no está ocurriendo aquí») en defensa de una «teoría A» del espacio? ¿Arrojaría ello dudas sobre el argumento inicial de Prior?
- ¿Es el mundo imaginario de Shoemaker realmente un mundo posible? En caso afirmativo, ¿constituye dicho mundo un problema para la teoría B?
- ¿Hay paradojas genuinas respecto de viajar en el tiempo?
- ¿Podría un viajero en el tiempo ser su propio padre?

### Notas

1. A. N. Prior, «Thank Goodness That's Over», *Philosophy,* vol. 34, 1959, p. 17.
2. H. Mellor, *Real Time* (Cambridge: Cambridge University Press) 1985, p. 48.
3. Véase D. Parfit, *Reasons and Persons* (Oxford: Oxford University Press) 1986, capítulo 8.

4. Esto no es lo mismo que preguntar si podría haber un mundo en el que el tiempo transcurriera sin que nada sucediese. Lo que preguntamos es si, en un mundo en el que ocurren muchas cosas, podría haber un periodo (por pequeño que fuera) en el que nada sucediese.
5. S. Shoemaker, «Time without Change», reimpreso en R. Le Poidevin y M. Macbeath (eds.), *The Philosophy of Time* (Oxford: Oxford University Press) 1993, pp. 63-79.
6. Aristóteles, *Física*, Libro IV, capítulo 11, 218b.
7. Shoemaker, *op. cit.*, p. 67.
8. Suponiendo, esto es, que la ley que gobierna el susodicho universo no sea que las paralizaciones ocurren en el sentido indicado excepto por el hecho de que «las tres regiones en conjunto se saltan una paralización cada cincuenta y nueve años» (Shoemaker, *op. cit.*, p. 72). Pero parece justificado creer en la ley más simple libre de excepciones.
9. Shoemaker, *op. cit.*, p. 75.
10. D. Lewis, «The Paradoxes of Time Travel», *American Philosophical Quarterly,* vol. 13, n.º 1, 1976, pp. 145-152.
11. Ibíd., p. 145.
12. Ibíd., p. 146. Hay posibilidades aun más exóticas: «Una cantidad infinita de tiempo personal puede ser comprimida en dos minutos de tiempo externo. Durante el primer minuto, el pseudoinmortal vive el primer día de su vida. Durante el siguiente medio minuto, el pseudoinmortal vive el segundo día. Durante el siguiente cuarto de minuto, transcurre un tercer día. Debido a que hay un número infinito de eslabones en esta secuencia, el pseudoinmortal disfrutará de un número infinito de días personales» (R. Sorenson, «The Cheated God: Death and Personal Time», *Analysis,* vol. 65, n.º 2, 2005, pp. 122-123).
13. Lewis, *op. cit.*, pp. 146-147.
14. Un caso que podría ser problemático sería el de alguien que viajase instantáneamente en el tiempo, por ejemplo, que viajase a 1900 pero no utilizase para ello tiempo alguno. La «línea entre mundos» de tal persona sería discontinua, en lugar de continua. Sin embargo, aun así cabría apelar a factores causales para justificar la afirmación de que la persona que partió en 2009 es la misma que la que llegó a su destino en 1900, y no un mero doble. Si las características mentales y físicas de la persona de 1900 fueron causadas por sus estados en 2009, entonces puede justificarse el hablar de una misma persona.
15. Ibíd., p. 148.
16. Ibíd., pp. 148-149.
17. Ibíd., p. 149. Para algunas reflexiones interesantes sobre los circuitos causales, véase R. Hanley, «No End in Sight: Causal Loops in Philosophy, Physics and Fiction», *Synthese,* vol. 141, 2004, pp. 123-152.
18. Ibíd., p. 149.
19. Ibíd., p. 149.
20. Ibíd., p. 150.

## Bibliografía adicional comentada

P. Horwich, *Asymmetries in Time* (Cambridge, Mass.: MIT Press) 1987, capítulo 7. Una defensa bastante técnica, pero todavía útil, de la tesis de Gödel de que viajar al pasado es psicológicamente imposible.

D. Lewis, «The Paradoxes of Time Travel», *American Philosophical Quarterly,* vol. 13, n.º 1, 1976. Reimpreso en sus *Philosophical Papers,* vol. II (Oxford: Oxford University Press) 1986, capítulo 18. El tratamiento filosófico clásico del viaje en el tiempo. El artículo de Lewis es muy ameno y está libre de tecnicismos.

H. Mellor, *Real Time* (Cambridge: Cambridge University Press) 1985. La defensa mejor sustentada que existe de la teoría B. Contiene una respuesta al argumento de Prior de «menos mal que eso acabó».

A. N. Prior, «Thank goodness that's over», *Philosophy,* vol. 34, 1959. Formulación clara de un problema de la teoría B que se ha convertido en clásico.

S. Shoemaker, «Time without Change», en R. Le Poidevin y M. MacBeath (eds.), *The Philosophy of Time* (Oxford: Oxford University Press) 1993. Una presentación clara y detallada de los argumentos a favor de la posibilidad de un tiempo sin cambio.

## Recursos en Internet

P. Horwich (1998), «Time travel», en E. Craig (ed.), *Routledge Encyclopedia of Philosophy.* Bajado el 31 de mayo de 2006 de <http://www.rep.routledge.com/article/Q108>.

F. Arntzenius y T. Maudlin (2005) «Time Travel and Modern Physics», *The Stanford Encyclopedia of Philosophy (edición de verano de 2005),* Edward N. Zalta (ed.). Bajado el 31 de mayo de 2006 de <http://plato.stanford.edu/archives/sum2005/entries/time-travel-phys>.

# 7. El libre albedrío

## Introducción

En este capítulo, nos ocuparemos de dos ataques a la tesis de que nuestra voluntad es libre, o tesis del libre albedrío. Podría parecer insólito que se ataque esta tesis, ya que creemos estar actuando libremente todo el tiempo. Sin embargo, el libre albedrío ha sido cuestionado desde los albores de la filosofía.

El primer ataque al libre albedrío consta de una serie de reflexiones y argumentos conocidos como fatalismo. Los argumentos del fatalista típico son de carácter puramente lógico o a priori: es decir, no descansan sobre premisas empíricas. A pesar de ello, otra versión del fatalismo —el fatalismo teológico— se basa en presupuestos sobre la existencia y la naturaleza divinas.

El segundo frente de ataque contra el libre albedrío reposa sobre una premisa empírica, proveniente del determinismo. De acuerdo con este segundo frente, nuestro universo es determinista y la verdad del determinismo es incompatible con el libre albedrío, razón por la cual lo excluye. La tesis del determinismo, en caso de ser cierta, sería una verdad contingente y empírica. Mientras que los argumentos fatalistas distaban de ser convincentes, el ataque proveniente del determinismo al libre albedrío resulta más inquietante. Dando un giro adicional a la discusión, Galen Strawson ha argumentado recientemente con fuerza que la noción del libre albedrío es incoherente y, por ende, no compatible ni con el determinismo ni con el indeterminismo.

El debate sobre el libre albedrío es importante no sólo como una cuestión de metafísica, sino también porque se suele dar por sentado que la responsabilidad

moral exige el libre albedrío; y que, por lo tanto, si el libre albedrío fuera una ilusión, también lo sería la responsabilidad moral, lo cual subvertiría las costumbres sociales, morales y legales que dan por sentada dicha responsabilidad.

## El fatalismo

La conclusión de la argumentación fatalista es, como su nombre indica, que somos prisioneros de la fatalidad: no podemos hacer sino lo que de hecho hacemos. La versión más sencilla del fatalismo se remite a dos principios a primera vista plausibles:

(i) Habrá un único futuro real.
(ii) Para cualquier proposición P, si P es verdadera ahora, fue verdadero que P en cualquier tiempo pasado.

El primer principio trata de captar la tesis incontrovertible de que no hay dos o más futuros reales. Aunque aún no la conozcamos, sólo hay una forma en que el universo evolucionará después de un tiempo dado. El segundo principio es una versión del principio de intemporalidad de la verdad: si, por ejemplo, ahora es cierto que Tony Blair ganó las elecciones británicas generales de 2005, entonces ya era cierto en cualquier tiempo pasado —digamos, en el año 900 a.C.— que Blair ganaría las elecciones de 2005.

¿Cómo podrían estos principios generar una conclusión fatalista? Supongamos que P sea un enunciado sobre un futuro contingente. Se refiere a algún tiempo en el futuro, y no es necesariamente verdadero. Además, P se refiere a una de nuestras acciones. En especial, sea P = Yo votaré por el partido liberal el año próximo. En este momento sólo hay un único futuro real (por el principio [i]). Supongamos que yo vote por el partido liberal el año próximo. Por lo tanto, es verdad ahora que yo votaré por el partido liberal el año próximo. Según el principio (ii), para cualquier tiempo pasado, digamos 1800, fue verdadero en ese tiempo pasado que yo votaría el año próximo por el partido liberal. Pero entonces, concluye el fatalista, es inevitable que yo vote por el partido liberal el año próximo. Soy incapaz de actuar de otra manera. Lo mismo se aplica a cualquier otra acción mía en el futuro que yo considere libre. La libertad sería, pues, una ilusión.

Así, de:

(1) Votaré por el partido liberal el año próximo, y
(1) En 1800 ya era verdadero que yo voy a votar por el partido liberal el año próximo,

el fatalista concluye,

(3) Es inevitable que yo vote por el partido liberal el año próximo.

## El fatalismo

Los fatalistas sostienen que ya era verdadero hace cien años que yo escribiría estas palabras hoy. Los fatalistas teológicos afirman que Dios siempre supo que yo escribiría estas palabras hoy. De ahí concluyen que era inevitable que yo escribiera estas palabras hoy. Yo no habría podido no hacerlo. Esta acción, como todas las acciones que lleva a cabo la gente, no sería, por lo tanto, libre. La libertad sería una ilusión. Por fortuna para nosotros, el razonamiento de los fatalistas es confuso. Nos quieren hacer creer que algún hecho sobre el pasado o sobre la presciencia de Dios hace necesario, o hace que sea verdadero, que yo escriba estas líneas hoy. Pero la intuición nos indica que la dirección de dependencia es la inversa. Era verdadero hace cien años que yo escribiría hoy estas líneas porque ahora las estoy escribiendo por voluntad propia. ¿No hace esto que un hecho referido al pasado dependa de un hecho referido al presente? No, porque el hecho supuestamente referido al pasado se refiere a él sólo de manera aparente y gramatical. En realidad, se trata de un hecho referido al presente, pero disfrazado como un hecho referido al pasado. De modo similar, si Dios sabía que yo habría de escribir estas palabras hoy, es porque yo he elegido libremente escribirlas hoy. El argumento del fatalista está, pues, viciado.

Las acciones futuras no son las únicas que carecen de libertad. Alterando los tiempos verbales, podemos aplicar el mismo razonamiento para mostrar que ninguna de mis acciones pasadas fue libre. Así, de:

(1*) Voté por el partido liberal el año pasado, y
(2*) En 1800 ya era verdadero que yo votaría por el partido liberal el año pasado,

el fatalista concluye

(3*) Fue inevitable que yo votara por el partido liberal el año pasado.

Lo mismo vale para cualquier acción pasada mía que yo creía haber ejecutado libremente. Por lo tanto, generalizando mi caso a los de otras personas, el fatalista concluye que ninguna acción de agente alguno ha sido, es o será libre.

Esta conclusión resulta escandalosa. ¿Hay alguna forma de evitarla? Davis Lewis ofrece el diagnóstico siguiente:

Los fatalistas —los mejores entre ellos— son filósofos que toman ciertos hechos que nosotros consideramos irrelevantes cuando especificamos lo que alguien puede hacer; luego los disfrazan como especies de hechos del tipo que solemos considerar relevantes; y terminan argumentando que podemos hacer menos de lo que creemos que podemos hacer; es más, que no hay nada, aparte de lo que ya hacemos, que podamos hacer [...] [Un hecho

como (2)] es un hecho irrelevante sobre el futuro, que es disfrazado como un hecho relevante sobre el pasado y que, por lo tanto, debería ser dejado de lado cuando se habla de lo que, en un sentido ordinario del término, yo puedo hacer[1].

Para establecer lo que alguien puede hacer ahora, son relevantes los hechos sobre el pasado. Así, el que yo pueda, o no, nadar ahora está determinado por hechos pasados como el de si yo tomé, o no, clases de natación. De ese modo, el hecho pasado:

(4) Nunca tomé clases de natación

es un factor relevante para el valor de verdad de:

(5) Puedo nadar ahora.

La verdad de (4) explica la falsedad de (5). El enunciado (4) es un hecho que está referido enteramente al pasado; (5) es un hecho sobre el presente. El fatalista considera la relación entre (2) y (1) análoga a la que se da entre (4) y (5). Así como el hecho pasado de que nunca tomé clases de natación explica el hecho y la razón de que yo no pueda nadar ahora, así, piensa el fatalista, el hecho pasado de que en 1800 ya fuera cierto que yo iba a votar por el partido liberal el año próximo explica el hecho y la razón de que yo no pueda dejar de votar por el partido liberal el año próximo. Pero, en eso, piensa Lewis, consiste el error del fatalista: (2) se parece a un hecho sobre el pasado, pero en realidad no lo es. Sería un hecho sobre el año próximo, que está disfrazado para que parezca un hecho sobre 1800.

El punto destacado por Lewis es plausible. No pensamos que (2) verse sobre 1800 en el mismo sentido en que, por ejemplo,

(6) En 1800 la población de Londres superaba el millón de personas

versa sobre 1800. A pesar de ello, algo parece estar faltando. Como Lewis no discute la verdad de (1) y (2), ¿por qué, exactamente, no estaríamos justificados en inferir (3)? ¿Qué pasaría si el fatalista contestase que, para él, (2) es relevante para evaluar la verdad de (1), y que de (1) y (2) se deduce (3)? ¿Cómo deberíamos responder?

Algunos comentaristas han señalado que el fatalista hace una presuposición peculiar sobre la dirección de dependencia entre (1) y (2)[2]. Para poder inferir (3) a partir de (1) y (2), el fatalista debe dar por sentado que el que yo vote por el partido liberal el año próximo depende del hecho de que en 1800 ya era cierto que yo iba a votar por el partido liberal el año próximo. Pero es obvio intuitivamente que la dirección de dependencia es precisamente la opuesta: el que sea verdadero en 1800 que yo vaya a votar por el partido liberal el año próximo depende del hecho de que votaré por el partido liberal el año próxi-

mo. Por supuesto, si (2) depende de (1), da la impresión de que un hecho pasado depende de un hecho futuro. Es aquí donde la observación de Lewis cobra relevancia: (2) no es realmente un hecho sobre el pasado; es un hecho sobre el futuro, pero disfrazado. Una vez que nos damos cuenta de que (2) depende de (1), y no al revés, y de que (2) no es un hecho sobre el pasado, no hay razón para pensar que (1) y (2) impliquen (3). Por lo tanto, tenemos todo el derecho a considerar inválido el argumento (1)-(3) del fatalista.

### El determinismo

Ha sido tradicional creer que la tesis del determinismo amenaza la libertad de acción y de pensamiento. Si el determinismo es verdadero, entonces, dadas las leyes de la naturaleza y el estado del universo en cualquier tiempo pretérito dado, estaba predeterminado que tú leyeras estas líneas ahora. Y como es físicamente imposible que tú no estés leyendo estas líneas, tu lectura de ellas no sería libre. Los compatibilistas sostienen que el determinismo no amenaza nuestra libertad. Los libertaristas, en cambio, afirman que sí la amenaza, y que, precisamente por esta razón, el determinismo es falso. Pero ¿no debería ser la ciencia la encargada de decirnos si el determinismo es verdadero o falso? Comoquiera que sea, ¿cómo puede el indeterminismo dotar al libre albedrío de un marco adecuado? Las acciones cuidadosamente razonadas y predecibles no parecen ser fruto de eventos indeterminados o parcialmente azarosos. Otros, finalmente, sostienen que tanto los compatibilistas como los libertaristas se equivocan. Según ellos, el libre albedrío es lógicamente imposible. Yo no podría ser «la raíz última de todas mis determinaciones» del modo en que lo exige el concepto de libre albedrío. Las atribuciones de responsabilidad moral, dado que presuponen el libre albedrío, carecerían de fundamento.

## Un argumento fatalista más sofisticado

¿Tiene el fatalista otro modo, quizás más sofisticado, de presentar su argumento? Considérese a continuación la forma en que Michael Dummett aborda la cuestión en su artículo «Creando el pasado». Escribe lo siguiente:

> La forma estándar del argumento del fatalista fue muy popular en Londres durante los bombardeos. Una sirena suena, y me dirijo a los refugios antiaéreos para evitar ser matado por una bomba. El fatalista arguye: «O te va a matar una bomba, o no te va a matar una bomba. Si te va a matar una bomba, cualquier precaución que tomes será inútil. Si no te va a matar una bomba, toda precaución será superflua. Por lo tanto, no tiene sentido que tomes precauciones»[3].

Dummett examina este argumento porque se da cuenta de que el principal argumento usado contra la racionalidad de tratar de modificar el pasado (por

ejemplo, con plegarias retrospectivas) es muy parecido al argumento del fatalista, excepto, por supuesto, por la modificación del tiempo verbal. Así, en respuesta a un padre que, al desconocer la suerte de su hijo, reza a Dios en el presente para que su hijo haya logrado salvarse de morir ahogado el día de ayer, Dummett imagina que se podría decir lo siguiente:

> O tu hijo se ahogó o no lo hizo. Si se ahogó, entonces es obvio que tu plegaria no va a ser (no puede ser) escuchada. Si no se ahogó, entonces tu plegaria es superflua. Por lo tanto, en cualquiera de los dos casos, tu plegaria no tiene sentido: no puede *influir* en la cuestión de si tu hijo se ahogó o no[4].

La idea de Dummett es que como no aceptamos el argumento del fatalista, tampoco deberíamos aceptar el presente argumento, lo cual nos permitiría al menos darle sentido a la idea de hacer algo ahora con la finalidad de que otra cosa haya ocurrido anteriormente (al menos cuando uno no sabe aún si el evento anterior ocurrió o no).

Por supuesto, se podría objetar que los dos argumentos no son análogos, precisamente a causa de la diferencia en el tiempo verbal. Dado que el pasado es real, el argumento en contra de las plegarias retrospectivas sería inválido. Podemos denominar a esto «la solución aristotélica»[5].

Para poder evaluar esta solución, desglosemos primero el argumento del fatalista:

(1) O bien te va a matar una bomba o no te va a matar una bomba.
(2) Si te va a matar una bomba, cualquier precaución que tomes será inútil.
(3) Si no te va a matar una bomba, cualquier precaución que tomes será superflua.

Por lo tanto:

(4) No tiene sentido que tomes precauciones.

De acuerdo con la solución aristotélica, los enunciados contingentes futuros carecen de valor de verdad. La solución aristotélica presupone la irrealidad del futuro: no hay realidad alguna que haga que los enunciados contingentes futuros sean verdaderos o falsos ahora, a diferencia de lo que sucede con los enunciados necesarios futuros, que sí son verdaderos o falsos ahora. Es natural que un aristotélico caracterice la irrealidad o «apertura» del futuro utilizando la figura de un árbol. En el presente, hay muchas ramas futuras posibles (todas compatibles con el estado actual de las cosas); a medida que el presente (o *ahora* en movimiento) se va desplazando, se desgajan todas las ramas excepto una[6]. (Véase el capítulo 5).

El aristotélico puede, a continuación, afirmar que un enunciado sobre el futuro es verdadero, si es verdadero en cada rama; falso, si es falso en cada rama, y ni verdadero ni falso, en cualquier otro caso. Así, un enunciado contingente futuro, como «lloverá mañana», no es ahora ni verdadero ni falso, ya que es verdadero en algunas ramas, pero falso en otras. En contraste, la verdad necesaria futura «lloverá o no lloverá mañana» es ahora verdadera, ya que es verdadera en cada rama; y la falsedad necesaria futura «lloverá y no lloverá mañana» es ahora falsa ya que es falsa en cada rama. Nótese que esta explicación es violatoria de la lógica clásica, ya que se está afirmando la verdad de una disyunción («lloverá o no lloverá mañana») ninguno de cuyos miembros es verdadero.

¿Qué es lo que la solución aristotélica encuentra errado en la argumentación del fatalista? No, ciertamente, la primera premisa. Pero aunque la premisa (1), que es una disyunción, sea verdadera, ninguno de sus miembros disyuntos es verdadero, y, por lo tanto, tampoco los antecedentes de (2) y (3) serán verdaderos. El aristotélico puede, entonces, sostener o bien que los **condicionales** con antecedentes carentes de valor de verdad son ellos mismos falsos (en cuyo caso la argumentación del fatalista contendrá dos premisas falsas); o bien que el razonamiento por disyunción (P o no-P; si P, entonces R; si no-P, entonces R; luego R) no es válido en los casos en que «P» y «no-P» carecen de valor de verdad (en cuyo caso la argumentación del fatalista será inválida). Por consiguiente, la argumentación del fatalista o bien será infundada, o bien será inválida. Sin embargo, de acuerdo con la solución aristotélica, no existe un fallo análoga en el argumento contra la plegaria retrospectiva: «él se ahogó» y «él no se ahogó» son ambos determinados en cuanto a su valor de verdad.

Pero hay un error más grave en la argumentación del fatalista. Incluso si los enunciados contingentes futuros tuvieran un valor de verdad, la premisa (3) distaría mucho de ser plausible. Es un sofisma argüir que si a ti no te van a matar, cualquier precaución que tomes será superflua. Podría suceder que no te mataran precisamente porque tomas precauciones.

Para que pudiera ser cierto que las precauciones han sido superfluas, tendría que ser verdadero el **condicional contrafáctico** siguiente: si no hubieras tomado precaución alguna, (aun así) no te habrían matado. Pero la verdad de este contrafáctico no se sigue únicamente de la verdad de «no te matarán». El principio «si P, entonces, si Q hubiera sido el caso, (aun así) P sería el caso» (para cualesquiera P y Q) es un principio obviamente falso. Supóngase que yo tuve un accidente automovilístico y me salvé porque llevaba puesto el cinturón de seguridad. Sería descabellado razonar: «no me maté, por lo que, si no hubiera llevado puesto el cinturón, (aun así) no me habría matado».

Lo que sí sigue de cualquier verdad P, al menos de acuerdo con ciertas interpretaciones del **condicional de indicativo**, es: si Q, entonces P. Es decir, «si P, entonces, si Q entonces P» es una tautología, suponiendo que «si... entonces» es entendido en castellano como el **condicional material**. Según esta interpretación de los condicionales en indicativo, lo siguiente es verdadero: si

no te matarán, entonces, si no tomas precauciones, no te matarán. Pero esto no implica que las precauciones hayan sido superfluas. Como hemos visto, la superfluidad requiere no la verdad de un condicional de indicativo, sino la verdad de un condicional contrafáctico, y ese contrafáctico no se sigue meramente de la verdad de «no te matarán».

Consecuentemente, la argumentación a favor de (3) es falaz, y deberíamos simplemente rechazar esta última premisa. La mera verdad (si es que es verdad) de «no te matarán» no es evidencia alguna de que cualquier precaución que hayas podido tomar fue superflua. Por esa misma razón, deberíamos rechazar el argumento anterior contra la utilidad de la plegaria retrospectiva. Pues la premisa «si él no se ha ahogado, tu plegaria es superflua» está expuesta exactamente a la misma objeción recién formulada contra la premisa (3).

Sin embargo, hay aparentemente circunstancias extraordinarias en las que una actitud fatalista hacia la propia vida parece justificada. He aquí una situación imaginaria muy vívida: Estás ojeando un día los estantes de una biblioteca presencial y te topas con un libro de apariencia extraña titulado *He aquí tu vida.* Al empezar a leerlo, te das cuenta de que se trata, efectivamente, de un libro sobre tu vida. Contiene una descripción exacta de tu educación, tu carrera, tu apariencia, tu comportamiento e, incluso, tus pensamientos más íntimos. Hay por escrito verdades sobre ti mismo que, por descontado, nadie más que tú habría podido conocer. El libro parece viejo, y su fecha de publicación es anterior a la de tu nacimiento. También te das cuenta de que el libro no concluye en el presente; antes bien, la descripción del presente figura apenas a eso de la mitad del libro. Impaciente, comienzas a leer sobre tu vida futura...

En estas condiciones, ¿no sería natural adoptar una actitud fatalista hacia la vida de uno? Quizás pienses que esa actitud fatalista se debe a que hay verdades sobre el curso futuro de tu vida, como asevera el fatalista. Pero hemos visto que el fatalismo es un truco lógico, por lo que la mera existencia de verdades sobre tu vida no puede justificar una actitud fatalista respecto de ella. Ahora bien, tú podrías pensar también que la única manera posible en que el autor del libro *conozca* todas esas verdades sobre ti es que él sea también el autor de tu vida, es decir, que él tenga el control de la situación y tú seas su marioneta. En ese caso, una actitud «fatalista» sí estaría justificada, ya que ninguna de tus acciones te pertenece y, por lo tanto, ninguna es libre. Pero lo que en ese caso justificaría tu actitud no sería la existencia de verdades sobre ti, sino el hecho de que tus ideas y actos son controlados por otra persona. Consecuentemente, el fatalismo, tal como es entendido aquí, no queda reivindicado por la situación imaginaria anteriormente descrita.

## El fatalismo teológico

Existe otra versión del fatalismo, objeto de muchos debates en la Edad Media, que surge de la supuesta tensión entre la omnisciencia divina y el libre albe-

drío humano. Si Dios conoce todas las verdades (pasadas, presentes y futuras), entonces sabe lo que hará cualquier ser humano. ¿Cómo, entonces, puede ser libre cualquier acción humana? Es obvio que este problema se presenta en un contexto teológico en el que se presume la existencia y la omnisciencia de Dios. Ahora bien, incluso en ese contexto, ¿se puede decir que el fatalismo teológico da lugar a algún nuevo problema?

A. J. Ayer cree que no:

Si el hecho de que alguien sepa lo que yo voy a hacer mañana no hace necesario (inevitable) que yo lo haga, entonces el hecho de que alguien sepa lo que voy a hacer no sólo mañana, sino todos los días de mi vida, tampoco podrá hacer que esas acciones sean necesarias [...] El que la persona a la que se atribuye el conocimiento previo sea concebida como humana o como divina tampoco marca una diferencia [...] Todas estas consideraciones son irrelevantes[7].

De acuerdo con Ayer, si el hecho de que ahora sea cierto que yo votaré por el partido liberal el año próximo no hace inevitable que lo haga, entonces el conocimiento de esa misma verdad (por parte de Dios o de un hombre) tampoco hace inevitable que yo realice esa acción. Ni siquiera el conocimiento de muchas más verdades acerca de mí cambiará la situación.

Hay versiones del fatalismo teológico que parecen introducir un elemento novedoso. Steven Cahn escribe lo siguiente:

Si es verdadero que yo realizaré una acción particular, entonces Dios, que conoce todas las verdades, sabe que yo realizaré esa acción. Pero si yo pudiera abstenerme de realizar esa acción, entonces, aparentemente, yo podría refutar el conocimiento divino, lo cual es imposible. Y si no puedo abstenerme de realizar la acción de marras, es porque ésta no es libre[8].

Pero ¿es este razonamiento realmente novedoso? ¿No podría haber sido presentado en un contexto no teológico? Verbigracia: «Si es verdadero ahora que yo realizaré una acción determinada A, entonces realizaré esa acción. Pero si yo pudiera abstenerme de realizar la acción A, tendría la capacidad de modificar el pasado, lo cual es imposible. Por otra parte, si yo no pudiera abstenerme de realizar esa acción, entonces la acción no sería libre».

La respuesta a este argumento es que no es verdad que si yo puedo abstenerme de realizar la acción A, ello significa que tengo la capacidad de modificar el pasado. En este mundo, realizo la acción A. En otro mundo posible diferente, en el que en lugar de A hago B, será verdadero en ese mundo que yo haré B. Se trata de dos mundos posibles que contienen verdades diferentes y acciones diferentes. En ninguno de ellos ha sido cambiado el pasado.

La misma respuesta se puede formular en clave teológica: no es verdad que si yo pudiera abstenerme de realizar la acción A, estaría refutando el conocimiento divino. En vez de ello, en el mundo en el que yo hago B, el conocimiento de Dios en ese mundo es diferente del conocimiento de Dios en este mundo

(ya que estos mundos contienen, respectivamente, verdades diferentes). En el mundo A, Dios sabe que haré A; en el mundo B, Dios sabe que haré B. Esto no refuta el conocimiento divino. Adicionalmente, el argumento teológico incurría en la suposición fatalista estándar, diagnosticada arriba, sobre la «dirección de dependencia». El conocimiento divino de que yo haré A depende de que yo haga libremente A, y no al revés, como suponía el fatalista teológico.

## El libre albedrío y el determinismo

El fatalismo no es la única amenaza que se cierne sobre el libre albedrío. Desde la época de Thomas Hobbes (1588-1679), a los filósofos les ha inquietado que el determinismo suponga un peligro para el libre albedrío. Algunos —los compatibilistas (como Hobbes y David Hume)— consideran que dicha tensión es ilusoria, y alegan que el libre albedrío y el determinismo son compatibles. (Incluso algunos compatibilistas sostienen que el libre albedrío requiere que sea verdadero el determinismo.) Otros —los incompatibilistas— afirman que dicha tensión es completamente real: el libre albedrío y el determinismo serían incompatibles. Sin embargo, los incompatibilistas se dividen luego en dos bandos: los libertaristas [*libertarians*], quienes concluyen que, puesto que somos libres, el determinismo debe ser falso; y los deterministas radicales [*hard determinists*], quienes concluyen que, puesto que el determinismo es verdadero, carecemos de libre albedrío.

### Thomas Hobbes (1588-1679)

Hobbes es conocido sobre todo como filósofo de la política. En su libro *Leviatán*, publicado en 1651, abogó por el gobierno de un soberano absoluto como baluarte frente a los desórdenes civiles. Hobbes quería demostrar que la política podía ser una ciencia. Su explicación de la naturaleza de la materia sirvió de fundamento a su teoría de la psicología humana, que a su vez proporcionó la base para su teoría política. Según Hobbes, el universo entero, incluyendo el pensamiento y la conducta humana, es sólo materia en movimiento. Toda causalidad es de tipo mecanicista y necesario: todo está determinado por lo que lo precede. No obstante, Hobbes creyó que su concepción mecanicista del universo era congruente con la existencia de Dios (concebido como un ser material), con la de los estados mentales conscientes (también concebidos como físicos) y con la libertad humana. Hobbes fue, por consiguiente, uno de los primeros compatibilistas: «La libertad y la necesidad son consistentes entre sí», escribió. «Las acciones que los hombres llevan a cabo voluntariamente [...] proceden de la libertad; y, como cada acto de la voluntad humana [...] tiene una causa determinada, y ésta, a su vez, otra causa, de manera que se forma una cadena ininterrumpida, [las acciones voluntarias] provienen de la necesidad» (*Leviatán*).

Finalmente, están aquellos filósofos (como Galen Strawson) que sostienen que la verdad o falsedad del determinismo es irrelevante para la cuestión de si tenemos o no libre albedrío. El concepto de libre albedrío —ya se trate del libre albedrío compatibilista o del libertario— sería internamente incoherente, y para mostrarlo bastarían razones a priori. Por lo tanto, el libre albedrío sería una ilusión. Como la responsabilidad moral presupone el libre albedrío, la responsabilidad moral sería también una ilusión[9].

## Defensa del incompatibilismo

Para poder apreciar lo que habla a favor del incompatibilismo necesitamos primero caracterizar al determinismo. El determinismo es la tesis de que, dadas las **leyes de la naturaleza** y el estado del universo en cualquier tiempo pasado *t,* es físicamente imposible que la historia del universo (antes y después de *t*) sea de otra forma que como realmente es. Otro modo de formular el determinismo es en términos de causalidad. Puede ser descrito como la tesis según la cual cada evento tiene una causa, entendiendo causas como algo que determina necesariamente sus efectos.

Tanto las regularidades que observamos a nuestro alrededor como el surgimiento de la ciencia newtoniana son factores que refuerzan la creencia en el determinismo. Pero la tesis del determinismo no es una verdad lógica. Incuso si fuera verdadera, su verdad sería contingente, empírica. No obstante, la tesis del determinismo puede ser criticada por medio de razones filosóficas. Por ejemplo, si aceptáramos una teoría de la causalidad como regularidad (del tipo asociado a Hume), tendríamos una razón para negar que las causas determinen necesariamente sus efectos. Pero, como ya vimos en el capítulo 4, la teoría de la regularidad está abierta a objeciones.

Por supuesto, muchos piensan que los principios de la mecánica cuántica revelan que nuestro universo no es determinista. Pero no está claro hasta qué punto los sucesos indeterministas, o parcialmente azarosos, del nivel subatómico podrían ser relevantes para la cuestión del libre albedrío. Si la acción de un agente es el resultado de un proceso indeterminista, ¿cómo explicar por qué realizamos esa acción en lugar de alguna otra, alternativa? ¿Cómo podría un agente ser moralmente responsable de una acción que es resultado de un proceso indeterminista (es decir, un proceso sobre el cual carecemos de control)? ¿Por qué tal agente, cuando hace algo correcto, no está siendo simplemente favorecido por la suerte? Cuestiones como éstas consituyen problemas considerables para las concepciones libertaristas del libre albedrío.

Sin embargo, aunque el determinismo fracase en el nivel subatómico, el determinismo en el nivel macro podría constituir una amenaza para la libertad de acción, por lo que se requiere analizar el argumento del incompatibilismo. ¿Por qué se considera que el libre albedrío es incompatible con el determinismo? Supongamos que el determinismo es verdadero, y que yo realizo

libremente la acción A (digamos, muevo mi brazo para saludar). De acuerdo con el incompatibilista, si la acción A es libre, yo habría podido evitar hacer A. Podría haber hecho otra cosa. Pero, según el determinismo, dadas las leyes de la naturaleza y dado el estado del universo en cualquier tiempo previo al momento en que realicé A, el futuro del universo no habría podido ser sino de la manera en que es actualmente. Por lo tanto, yo no habría podido evitar hacer A. En ese caso, no hice libremente A; y lo mismo se aplica a cualquier otra acción o cualquier otra persona del universo.

A continuación mencionaré dos respuestas frecuentemente usadas contra este argumento incompatibilista. La primera respuesta alega que el argumento es únicamente válido si se presupone que las personas y sus estados mentales son puramente físicos. Pero no es verdad que el argumento se sostenga únicamente en estos términos. La mayoría de las acciones, como la de mover mi brazo, involucran el movimiento material, y tal movimiento está gobernado por leyes deterministas. Por ende, incluso si las personas y sus estados mentales fueran inmateriales, y aunque lo inmaterial no cayera bajo el control del determinismo, mi brazo no habría podido ocupar cualquier otro lugar espaciotemporal que el que realmente ocupó cuando lo moví. Así pues, incluso si las causas mentales del movimiento de mi brazo fueran inmateriales, seguiría siendo cierto que yo no habría podido abstenerme de mover mi brazo.

En segundo lugar, algunos filósofos han pensado que las acciones humanas no están regidas por leyes causales, ya que pertenecen a una esfera diferente. Las acciones humanas, a diferencia de los meros movimientos corporales, se explicarían mediante razones, y las razones no serían causas. Estaríamos operando con dos retículas explicativas distintas: el movimiento material se explicaría por medio de la ley causal, y la acción humana, por medio de la adscripción de razones, en cuyo caso la verdad o no verdad del determinismo sería irrelevante con respecto a la existencia y a la comprensión de la acción humana libre.

No obstante, es problemático aseverar que las razones no son causas. ¿Acaso mi deseo de saludar a alguien (es decir, una razón) no fue la causa de que yo moviera mi brazo? Pero hay un problema aún más elemental con esta respuesta, el mismo que aquejó a la respuesta anterior. Incluso si las acciones humanas y los movimientos corporales pertenecieran a categorías lógicas distintas, unas y otros permanecerían conectados entre sí: la mayoría de las acciones humanas, tales como agitar un brazo, implican movimientos corporales. Por ello, cualesquiera restricciones deterministas que haya del movimiento corporal serán también restricciones de la acción humana. En conclusión, ni la línea de argumentación «los estados mentales son inmateriales» ni la de «las razones no son causas» representan una réplica convincente al argumento del incompatibilismo.

## Una réplica compatibilista

¿Dónde, pues, piensan los compatibilistas que reside el fallo del argumento del incompatibilista? La respuesta compatibilista estándar consiste en negar que haya incompatibilidad alguna entre el libre albedrío y el determinismo. El libre albedrío debería ser contrastado no con la causalidad universal o el determinismo, sino con varias clases conocidas de *restricciones* [*constraints*]. Los compatibilistas, cuando dicen esto, suponen que su concepción está confirmada por el modo fáctico en que son usadas de ordinario las palabras «libre» y «no libre». Decimos que un hombre no es libre cuando, por ejemplo, alguien sostiene una pistola contra su sien, o lo encadena con grilletes, o cosas por el estilo. Es la presencia de tales restricciones específicas (coacciones) lo que priva a un hombre de su libertad, y no el hecho de la causalidad universal.

Sin embargo, hay dos razones para estar insatisfecho con esta respuesta. En primer lugar, el uso ordinario no es exactamente como lo describe el compatibilista. Si alguien sostiene una pistola contra mi sien, y, a consecuencia a ello, yo abro la caja fuerte, mi acción sigue siendo libre. Habría podido optar por no ceder a los deseos del ladrón (por muy insensato que ello hubiera sido). La tentación de describir este caso como un caso de acción no libre quizás resulte de confundir el libre albedrío con la responsabilidad moral. Es indudable que nadie me haría moralmente responsable por el robo del dinero en las circunstancias aludidas, pero eso no significa que mi acción no haya sido libre. Esta objeción no es, ciertamente, contundente contra el compatibilismo, pero sí demuestra que los casos de coacción con los que se debería contrastar la acción libre son aquellos (como la coacción física, la cleptomanía, la acción realizada bajo hipnosis, etc.) en los que uno está literalmente privado del poder de elegir o actuar.

En segundo lugar, y de manera más importante, el compatibilista está tratando de trazar una división, en el seno del ámbito de las causas, entre aquellas que son compatibles con el libre albedrío (es decir, las causas determinantes) y aquellas que no lo son (es decir, las causas coactivas). Pero ¿acaso no es arbitraria esta distinción? Cualquiera de estos dos tipos de causa impedirá que un agente actúe de otra manera que como lo hace; y ¿por qué habrían de ser compatibles con la acción libre las causas determinantes, pero no las coactivas?

Un experimento mental ideado por Harry Frankfurt puede quizás proveernos de una respuesta mejor al argumento del incompatibilista y, de esa manera, apuntalar una nueva concepción (compatibilista) del libre albedrío[10]. He aquí una versión simplificada del experimento mental de Frankfurt. Supongamos que Smith planea asaltar una sucursal de un banco, espera hasta la hora apropiada y ejecuta minuciosamente su plan. En términos de lo ocurrido, este asalto se parece a cualquier otro, salvo por un detalle. Sin que Smith lo sepa, un duende perverso, aficionado a los asaltos, ha estado monitorizando la mente de Smith, siempre dispuesto a intervenir en caso de que Smith flaquee en su empeño de realizar el asalto. Si Smith hubiese intentado cambiar de opi-

nión, se habría dado cuenta de que no podía. De hecho, sin embargo, Smith deseaba ansiosamente obtener el dinero y jamás cambió de opinión, por lo que el duende no tuvo que intervenir en el transcurrir de su vida.

Tenemos la creencia intuitiva sólida de que la acción de Smith era libre cuando éste estaba robando el banco, y que Smith es moralmente responsable de haberla realizado. La planificación y la ejecución de la acción corrieron completamente por cuenta suya. El duende no fue la causa de nada de lo que hizo. Y, sin embargo, Smith no podía hacer sino lo que hizo: con respecto al asalto, no habría podido elegir ni actuar de manera diferente de como lo hizo. ¿No demuestra esto que el libre albedrío no requiere de la capacidad para actuar de otra manera? En ese caso, el compatibilista dispone ahora de una réplica razonada frente al argumento del incompatibilista. La razón de que no se haya podido demostrar que el libre albedrío y el determinismo son incompatibles es que el libre albedrío no presupone realmente la capacidad de actuar de otra manera. Al contrario, lo único que requiere es que actuemos de acuerdo con nuestras creencias y deseos más propios[11].

## ¿Es incoherente la noción de libre albedrío?

A pesar de lo dicho, el ejemplo de Frankfurt no prueba que el compatibilismo sea cierto. Incluso si el determinismo no constituye una amenaza para el libre albedrío, el libre albedrío y el determinismo se mostrarán mutuamente incompatibles si el libre albedrío se revela algo imposible[12]. Galen Strawson ha argumentado recientemente que la noción de libre albedrío es incoherente[13]. El libre albedrío y la responsabilidad moral presuponen, según él, una concepción de autodeterminación lógicamente imposible de satisfacer. J. G. Fichte (1762-1814) supo captar agudamente esta paradójica noción de autodeterminación:

> Lo que yo deseaba era esto: que yo mismo, aquello que conscientemente concibo como mi propio ser y mi propia persona, [...] que este «yo» fuera independiente, fuera algo que existe no por otro y a través de otro, sino para sí mismo y que, como tal, fuera la raíz última de todas mis determinaciones[14].

¿Por qué habría de ser imposible la autodeterminación? La idea de Strawson es que la autodeterminación requiere que tú seas el responsable moral último de lo que haces. Pero para que tú seas el responsable moral último de lo que haces, debes ser el responsable último de lo que eres en ciertos aspectos mentales (como, por ejemplo, el carácter). Ahora bien, es imposible ser el responsable último de la manera en que eres en aspectos mentales (o en cualquier otro aspecto). Pues ello exigiría que tú hubieses causado intencionadamente tu posesión de cierta naturaleza mental. Y esto a su vez exigiría que tú tuvieras una naturaleza mental previa que hubieses causado intencionadamente, lo

cual a su vez exigiría una naturaleza mental previa que..., y así sucesivamente, *ad infinitum.* Por lo tanto, so pena de incurrir en un regreso al infinito, es imposible ser el responsable último de la forma en que eres en cualquier aspecto mental y, por lo tanto, es imposible ser el responsable último de lo que haces. A falta de tal responsabilidad, no puede haber autodeterminación ni, por consiguiente, libertad[15].

Este argumento es todo un reto. Se trata de un argumento sólido si cualquier noción cabal de libre albedrío o de responsabilidad moral ha de tener el mencionado carácter de autodeterminación o creación de sí mismo. El desafío, por lo tanto, consiste en determinar si el defensor del libre albedrío puede proporcionar o no la noción de un libre albedrío que no requiera una autodeterminación completa, a pesar de que ésta sirve de fundamento a la responsabilidad moral. Y esto podría revelarse un reto difícil de satisfacer.

## Galen Strawson (1952)

Galen Strawson, hijo del ya fallecido filósofo de Oxford P. F. Strawson, es actualmente catedrático de Filosofía en el programa de estudios de postgrado de la Universidad de la Ciudad de Nueva York. Recibió su formación en Cambridge, Oxford, y en La Sorbona, y ha sido además profesor en Oxford y Reading. Pensador original y riguroso, Strawson ha publicado hasta la fecha tres libros importantes. En 1986 apareció *Libertad y creencia,* en el que afirmaba que el libre albedrío y la responsabilidad moral radical son nociones incoherentes. El libre albedrío no era, según él, sino una ilusión. Luego vino, en 1989, *La conexión secreta,* obra en la que alegaba que Hume no sostuvo una teoría de la causalidad como regularidad pero sí creyó en la necesidad natural. En 1995 dio a la luz *La realidad mental,* donde aducía que la experiencia consciente es el signo distintivo de lo mental, y que no hay conexión conceptual alguna entre la posesión de estados mentales por parte de un sujeto y su comportamiento y sus disposiciones de carácter. Bautizó su posición como un «cartesianismo naturalizado». En estas tres obras Strawson combatió ingeniosamente dogmas bien asentados.

## Observaciones finales

Hemos examinado los dos ataques al libre albedrío más conocidos: el fatalismo y el determinismo. Los argumentos del fatalista fueron desenmascarados como sofismas. No lograban demostrar que lo que hacemos es inevitable o está fijado por el destino. El argumento incompatibilista clásico (que intenta demostrar que el libre albedrío y el determinismo son incompatibles) también se reveló fallido. La mayor amenaza contra el libre albedrío podría proceder de la tesis de Galen Strawson según la cual nuestras nociones de libre albedrío y responsabilidad moral última son lógicamente imposibles de satisfacer.

## Preguntas para el estudio

- Describa en líneas generales los dos argumentos fatalistas criticados en este capítulo. ¿Es mejorable alguno de ellos?
- Si rechazamos los argumentos del fatalista, ¿deberíamos aceptar también que a veces es racional intentar cambiar el pasado?
- ¿Hay algún aspecto interesante en el que el argumento del fatalista teológico difiera del de su contraparte no teológica?
- ¿Cómo defenderías la posición libertarista sobre el libre albedrío?
- ¿Hay alguna forma de justificar nuestras costumbres sociales y normas morales (como dar las gracias, censurar, castigar, etc.) sin apelar a la noción (supuestamente) incoherente de responsabilidad moral última?

### Notas

1. D. Lewis, «The Paradoxes of Time Travel», *American Philosophical Quarterly,* vol. 13, n.º 1, 1976, p. 151.
2. Por ejemplo, E. Craig, «Fatalism», en E. Craig (ed.), *Routledge Encyclopedia of Philosophy* (Londres: Routledge) 1998.
3. M. Dummett, «Bringing about the Past», en su obra *Truth and Other Enigmas* (Cambridge, Mass.: Harvard University Press) 1978, p. 339.
4. Ibíd., p. 338. Este argumento se utiliza para demostrar que la plegaria retrospectiva sería irracional aunque Dios existiese y la causalidad retrospectiva fuese posible.
5. Llamada así por la famosa discusión de Aristóteles en *De interpretatione,* ed. de C. W. A. Witaker (Oxford: Oxford University Press) 2002, cap. 9, 18a-19b.
6. Al presuponer el «carácter abierto del futuro», la solución aristotélica presupone también la falsedad del determinismo.
7. A. J. Ayer, «Fatalism», en su *The Concept of a Person* (Londres: Macmillan) 1963, pp. 252-253.
8. S. Cahn, «Fatalism», en J. Kim y E. Sosa (eds.), *A Companion to Metaphysics* (Oxford: Basil Blackwell) 1995, p. 169.
9. Algunos piensan que el problema más acuciante es el que se refiere a las condiciones de la responsabilidad moral. Según un punto de vista, primero decidimos en qué consiste que un agente sea moralmente responsable y luego elaboramos una noción de libre albedrío cónsona con ello. En su ensayo pionero «Freedom and Resentment» (en *Freedom and Resentment and Other Essays,* Londres: Methuen, 1974), P. F. Strawson se centra en nuestra práctica de atribuir responsabilidad a las personas, así como en prácticas similares, y concluye que éstas no apelan a una variable externa (como el determinismo o el indeterminismo) para legitimarse.
10. H. Frankfurt, «Moral Responsibility and Alternate Possibilities», *Journal of Philosophy,* vol. 66, 1969, pp. 829-839.
11. Por supuesto, la historia no acaba aquí. Se podría preguntar si la concepción compatibilista del libre albedrío es realmente una concepción del libre albedrío. El compatibilista concede que mis creencias y deseos actuales fueron causados (determinados) por eventos previos, los cuales a su vez habrían sido causados por eventos anteriores a ellos, y así sucesivamente, hasta llegar a eventos ocurridos antes

de que yo naciera. Entonces ¿cómo concebir «mis» creencias y deseos como propios?

12. Si el libre albedrío es imposible, no es compatible con nada (y, *a fortiori,* no es compatible con el determinismo).
13. G. Strawson, *Freedom and Belief* (Oxford: Oxford University Press) 1986. Véase también su artículo «The Impossibility of Moral Responsibility», *Philosophical Studies,* vol. 75, 1994, pp. 5-24.
14. J. G. Fichte, *The Vocation of Man,* ed. de R. Chisholm (Nueva York: The Liberal Arts Press) 1956, p. 27.
15. Nótese que la cuestión de la verdad o la falsedad del determinismo no desempeña papel alguno en este argumento.

## Bibliografía adicional comentada

A. J. Ayer, «Fatalism», en su obra *The Concept of a Person* (Londres: Macmillan) 1963. Una crítica clara, carente de sinsentidos, del fatalismo estándar y del fatalismo teológico.

—, «Freedom and Necessity», en sus *Philosophical Essays* (Londres: Macmillan) 1965. Una moderna defensa del compatibilismo sobre el libre albedrío.

R. Clarke, *Libertarian Accounts of Free Will* (Oxford: Oxford University Press) 2003. Un estudio panorámico que abarca los principales intentos libertaristas por dar sentido a la noción de libre albedrío.

M. Dummett, «Bringing about the Past», en su obra *Truth and Other Enigmas* (Cambridge, Mass.: Harvard University Press) 1978.Una discusión clásica, pero difícil, de la pregunta de si alguna vez puede ser racional proponerse modificar el pasado, y, de paso, una comparación con el argumento del fatalista y un diagnóstico de él.

G. Strawson, *Freedom and Belief* (Oxford: Oxford University Press) 1986. Defensa diáfana y amena de la concepción radical según la cual el libre albedrío y la responsabilidad moral última son ilusorios.

P. van Inwagen, *An Essay on Free Will* (Oxford: Clarendon Press) 1983. Alegato claro y exhaustivo a favor del incompatibilismo, con una discusión útil sobre el fatalismo.

G. Watson, *Free Will* (Oxford: Oxford University Press) 2003. Una valiosa colección de los mejores trabajos sobre el libre albedrío durante el último medio siglo.

## Recursos en Internet

E. Craig (1998) «Fatalism», en E. Craig (ed.), *Routledge Encyclopedia of Philosophy.* Bajado el 31 de mayo de 2006 de <http://www.rep.routledge.com/article/N096>.

M. McKenna (2004) «Compatibilism», *The Stanford Encyclopedia of Philosophy (edición de verano de 2004),* Edward N. Zalta (ed.). Bajado el 31 de mayo de 2006 de <http://plato.stanford.edu/archives/sum2004/entries/compatibilism>.

T. O'Connor (2005) «Free Will», *The Stanford Encyclopedia of Philosophy (edición de verano de 2005),* Edward N. Zalta (ed.). Bajado el 31 de mayo de 2006 de <http://plato.stanford.edu/archives/sum2005/entries/freewill>.

H. Rice (2002) «Fatalism», *The Stanford Encyclopedia of Philosophy (edición de invierno de 2002),* Edward N. Zalta (ed.). Bajado el 31 de mayo de 2006 de <http://plato.stanford.edu/archives/win2002/entries/fatalism>.

G. Strawson (1998, 2004) «Free Will», en E. Craig (ed.), *Routledge Encyclopedia of Philosophy.* Bajado el 31 de mayo de 2006 de <http://www.rep.routledge.com/article/V014>.

L. Zagzebski (2005) «Foreknowledge and Free Will», *The Stanford Encyclopedia of Philosophy (edición de otoño de 2005),* Edward N. Zalta (ed.). Bajado el 31 de mayo de 2006 de <http://plato.stanford.edu/archives/fall2005/entries/free-will-foreknowledge>.

# 8. La identidad personal

## Introducción

¿Qué es una persona? ¿Somos personas a lo largo de toda nuestra existencia? ¿En qué consiste para una persona persistir en el tiempo? ¿Es la identidad personal una relación importante? Éstas son las preguntas clave que serán abordadas en este capítulo.

Cuando formulamos preguntas sobre la identidad personal, nos referimos a «identidad» en el sentido de estricta identidad numérica, entendida de acuerdo con la **Ley de Leibniz** (es decir, A es idéntica a B sólo si todo lo que es verdadero respecto de A es verdadero respecto de B, y viceversa). No nos incumbe aquí la identidad en sentido cualitativo, como cuando hablamos de «gemelos idénticos». Desde un punto de vista numérico, los gemelos son dos personas, no una.

Nótese también que cuando preguntamos qué es para una persona persistir en el tiempo, estamos preguntando una cuestión constitutiva, no una cuestión evidencial. Estamos preguntando en qué consiste para una persona A en el tiempo $t_1$ ser la misma persona que la persona B en el tiempo $t_2$, y confiamos en que esta pregunta reciba una respuesta informativa. No estamos preguntando por qué creemos que A y B son idénticos, o en qué nos basamos normalmente para formarnos juicios sobre identidad personal. Esa evidencia (por ejemplo, la apariencia física, la voz, las huellas digitales, etc.) nunca podría fungir como garantía lógica de identidad personal. (Sería posible que alguien copiara mis huellas digitales, pero eso no haría que esa persona se convirtiera

en mí.) Lo que buscamos es una condición tal que, de cumplirse, garantice lógicamente la identidad; y que, de no cumplirse, garantice la no identidad. Aunque diferentes, la pregunta constitutiva y la evidencial están mutuamente relacionadas: se busca que la respuesta que se dé a la pregunta constitutiva concuerde con las maneras sencillas y convencionales de determinar quién es quién en los casos de la vida cotidiana.

### Criterios de identidad

Un criterio de identidad para los F nos indica en qué consiste la identidad de los F a través del tiempo, y, por lo tanto, nos dice qué cambios podría soportar F y qué cambios lo destruirían. Se da por sentado que el criterio de identidad para los F no presupone la noción de identidad de F. Por ello, se suele atribuir carácter reductivo a los criterios de identidad. Algunos creen que el criterio de identidad de artefactos, como los barcos, consiste en la continuidad espaciotemporal: un barco en $t_1$ es idéntico a un barco en $t_2$ si, y sólo si, el primer barco posee una continuidad espaciotemporal (sigue un curso ininterrumpido a través del tiempo y del espacio) con el segundo barco. El criterio es reductivo —pues la continuidad espaciotemporal no presupone la identidad de un barco— y nos dice a qué cambios puede sobrevivir un barco (esto es, qué cambios, aún por especificar, no interfieren en su continuidad espaciotemporal) y a cuáles no. Aunque el criterio está sujeto a objeciones —¿no podría acaso un barco sobrevivir a la discontinuidad (por ejemplo, si es desmantelado y reensamblado de nuevo)?—, su finalidad subyacente es bastante clara. ¿Existe un criterio para la identidad de las personas? Muchos responden que sí, pero los defensores de la concepción simple piensan que no: para ellos, la identidad personal es algo primitivo y básico. Las personas serían especiales precisamente porque carecen de criterios de identidad.

## Teorías de la identidad personal

Existen numerosas teorías de la identidad personal a través del tiempo. Es útil dividirlas en dos categorías: la concepción simple y la concepción compleja (esta última, en distintas variantes). Según la concepción simple, suscrita por Joseph Butler, Thomas Reid y, más recientemente, por Richard Swinburne, no es posible dar una respuesta sustancial e informativa a la pregunta constitutiva[1]. La identidad personal sería primitiva e inanalizable. No podemos especificar en otros términos lo que significa para la misma persona persistir a lo largo del tiempo. La concepción simple podrá parecer poco atractiva, pero no carece de motivos.

Según la concepción compleja, suscrita por la mayoría de los filósofos contemporáneos, la identidad personal es susceptible de un análisis ulterior,

por lo general en términos de continuidades de algún tipo. Algunas versiones de la concepción compleja analizan la identidad personal en términos de continuidad psicológica, mientras que otras versiones apelan a continuidades materiales, tales como la existencia continua del cuerpo o del cerebro.

Una distinción adicional puede ser trazada entre aquellas versiones de la concepción compleja que ofrecen un análisis **reductivo** de la identidad personal y aquellas que ofrecen meramente un análisis no reductivo. Un análisis de

## Joseph Butler (1692-1752)

Joseph Butler nació en Wantage, Inglaterra, y recibió su formación universitaria en Oriel College, Oxford, donde estudió las ideas de Locke y Shaftesbury. Sostuvo una correspondencia epistolar con el filósofo y teólogo inglés Samuel Clarke, y desarrolló su carrera en el marco de la iglesia anglicana, en la que sería eventualmente nombrado obispo de Durham. Sus libros principales son *Quince sermones* (1726) y *La analogía de la religión* (1736). En sus primeros escritos, Butler criticó a Shaftesbury por excluir de su teoría de la moralidad a la conciencia, y proporcionó una refutación del hedonismo psicológico o egoísmo. En su obra posterior, Butler argumentaría que es razonable aceptar argumentos a favor de la religión natural o deísmo (es decir, de los sistemas de pensamiento que procuran demostrar la existencia de Dios sobre la base de la razón, independientemente de cualquier revelación).

## Derek Parfit (1942)

Formado en Eton y Balliol College, Oxford, Parfit fue elegido Prize Fellow de All Souls College, Oxford, en 1967. Ha permanecido allí desde entonces, pero periódicamente visita las universidades de Harvard y Nueva York. Publicó una serie de artículos pioneros, comenzando con «Identidad personal» (1971); pero su obra más sobresaliente hasta la fecha ha sido su libro *Razones y personas* (1984). Gran parte de su abundante e ingeniosa obra está dedicada a criticar la teoría de la racionalidad como interés en uno mismo. Parfit ofrece numerosos argumentos en contra de esta teoría, y llega a la conclusión de que, por ejemplo, desear el beneficio de los demás a costa de uno mismo, o aspirar al progreso intelectual, no es menos racional que buscar el interés propio. Los puntos de vista de Parfit sobre la identidad personal también atacan la teoría del egoísmo. Si, como él dice, la identidad no es lo que importa, entonces la teoría del egoísmo pierde toda su fuerza. En la parte final de su libro, Parfit discute las perplejidades y paradojas surgidas de nuestra capacidad para influir tanto en la identidad de generaciones futuras como en la calidad de la vida de éstas. Necesitamos, dice, una nueva teoría de la benevolencia que dé una explicación adecuada de estos asuntos, pero reconoce que aún no la hemos encontrado.

la identidad personal es reductivo si el hecho de la existencia e identidad continuas de una persona puede ser comprendido sin hacer referencia al concepto de persona. Derek Parfit es el defensor más conocido del reduccionismo acerca de las personas. Piensa que la realidad puede ser completamente descrita en términos impersonales. Parfit tiene una concepción radical no sólo acerca de la naturaleza de la identidad personal, sino también acerca de su importancia. Sostiene que la identidad personal es trivial —no es la identidad lo determinante—, y cree que esto tiene implicaciones significativas para la racionalidad y la moralidad[2].

## ¿Qué es una persona?

Una persona es un ser que tiene actividad mental [*mental being*]. Pero no todos los seres que tienen actividad mental son personas: mi gato no lo es. Una persona es un tipo especial de ser que tiene actividad mental: un ser que tiene actividad mental y está dotado de autoconsciencia. Es difícil superar la definición dada por Locke, que sirve también para elucidar la noción de autoconciencia. Según Locke, una persona es «un ser pensante, inteligente, con raciocinio y capacidad reflexiva, que puede concebirse a sí mismo como un yo, una misma cosa pensante en tiempos y lugares diferentes»[3]. Así pues, una persona es un ser de cierta complejidad psicológica, capaz de concebir pensamientos temporalmente contextualizados [*tensed thoughts*] y condicionales contrafácticos en primera persona (como, por ejemplo, «hoy almorzaré en el Tetsuya», «fui al cine la semana pasada», «habría podido cenar anoche en el Oriental, pero decidí no hacerlo», etc.).

Esta definición satisface la condición deseada de incluir como personas a típicos seres humanos adultos, y permite asimismo apreciar por qué valoramos más a las personas que a los seres que no son personas, ya que la vida de un ser autoconsciente tiene más valor que la de un ser desprovisto de autoconsciencia (lo cual no significa que los seres no autoconscientes carezcan de valor). Sin embargo, es justo decir que esta definición, a pesar de ser verdadera en todo lo que expresa, no nos dice a qué categoría ontológica, o categoría del ser, pertenecen las personas. ¿Son las personas almas inmateriales, o seres humanos, o cuerpos, o cerebros, o conglomerados de percepciones?

Descartes creía que cada uno de nosotros tiene un alma inmaterial que sobrevive a la muerte corporal[4]. Hume, que las personas no son sustancias, sino conglomerados de percepciones, y su comparación del sujeto con una república se hizo célebre. Escribió: «No se me ocurre nada más apropiado con lo que comparar el alma que una república o mancomunidad cuyos distintos miembros, unidos por lazos recíprocos de gobierno y subordinación, generan a otras personas, que a su vez propagan esa misma república a través de los cambios incesantes que ésta experimenta»[5].

Parfit ha hecho mucho hincapié en esta analogía. Así como los elementos de una república (sus ciudadanos y su territorio) pueden ser comprendi-

dos sin hacer referencia al concepto de república, así los constituyentes del yo o de la persona (los pensamientos y las experiencias) pueden ser entendidos sin hacer referencia al concepto de persona. La analogía de Hume proporciona, por lo tanto, un modelo para el reduccionismo de Parfit respecto de las personas.

Hay, finalmente, concepciones diferentes según las cuales se nos puede identificar con nuestros cuerpos, o con nuestros cerebros, o con seres humanos (en el caso de que se entienda a los seres humanos como diferentes de los cuerpos humanos). Es amplio, pues, el espectro de puntos de vista acerca de cuáles son los auténticos portadores de autoconsciencia.

Aunque somos personas, no debería partirse del supuesto de que somos esencialmente personas, o que «persona» es, para usar la terminología de David Wiggins, un sortal de sustancia [*substance sortal*] en lugar de un sortal de fase [*phase sortal*][6]. En efecto, de acuerdo con una teoría reciente (el animalismo), aunque pueda haber personas no animales (por ejemplo, biónicas o inmateriales), somos esencialmente animales (seres humanos), pero no somos esencialmente personas[7]. Según esta teoría, soy idéntico al ser humano que está en mis zapatos. Como ese animal puede existir aunque se encuentre en un estado vegetativo permanente, irreversiblemente desprovisto de actividad mental, yo podría existir en ese estado, aunque una persona no podría hacerlo. En tal caso, yo existiría, aunque no como persona, y, por lo tanto, «persona» no puede ser un sortal sustancial. Según esta teoría, y contrariamente a lo que se ha venido suponiendo en la tradición, la cuestión de *nuestra* identidad no es la misma que la cuestión de la identidad *personal,* por lo que debería extrañarnos que a menudo se hayan planteado las preguntas equivocadas.

¿Cómo decidir entre estos modos de concebir la naturaleza de las personas? Varias estrategias están a nuestro alcance. Las discusiones sobre el dualismo de Descartes o sobre la teoría de Hume del sujeto tienen un carácter sumamente teórico. La decisión respecto de otras concepciones depende más, en cambio, de una metodología que recurre a los experimentos mentales. Aquí el hilo conductor es inequívoco: si una teoría sostiene que las personas son esencialmente F, y podemos imaginar un escenario en el que una persona continúa existiendo aunque no sea un F, entonces está claro que las personas no pueden ser idénticas a F, ni esencialmente F. En consecuencia, la mejor manera de responder a la pregunta «¿qué es una persona?» es responder primero a la pregunta «¿en qué consiste para una persona el persistir en el tiempo?».

## ¿En qué consiste para una persona persistir en el tiempo?

A continuación nos proponemos formular de manera más precisa las distintas versiones de la concepción compleja y examinar a qué problemas se enfrentan. Luego discutimos la tesis de Parfit de que la identidad personal no es lo

que importa, así como su reduccionismo sobre las personas. Finalmente, dirigimos nuestra atención a la concepción simple de la identidad.

Para comenzar, he aquí tres versiones de la concepción compleja, todas las cuales se ajustan a la mayoría de los juicios que emitimos sobre la identidad personal a lo largo del tiempo:

(1) El criterio corporal: A es en $t_1$ la misma persona que B en $t_2$, si, y sólo si, el cuerpo de A es el mismo que el cuerpo de B.
(2) El criterio cerebral: A es en $t_1$ la misma persona que B en $t_2$, si, y sólo si, el cerebro de A es el mismo que el cerebro de B.
(3) El criterio psicológico: A es en $t_1$ la misma persona que B en $t_2$, si, y sólo si, A y B son psicológicamente continuos.

Los criterios corporal y cerebral son aceptablemente claros. Se parte del supuesto de que la identidad de los cuerpos y de los cerebros a través del tiempo no es problemática, y que consiste en una continuidad espaciotemporal[8]. También se presupone, para poder generar un contraste nítido entre ambos criterios, que «cuerpo» significa «cuerpo menos cerebro». Aunque en el lado derecho de estos dos criterios aparecen los nombres de las personas («A» y «B»), es razonable presumir que estas menciones podrían ser eliminadas. El cuerpo de A, por ejemplo, podría ser especificado simplemente usando coordenadas espaciotemporales, sin hacer mención de A. Así pues, los criterios corporal y cerebral brindan análisis reductivos de la identidad personal. Una relación problemática —la de identidad personal— estaría siendo reducida a una relación que no lo es.

El criterio psicológico exige una explicación más prolija. Podemos decir que A y B son psicológicamente continuos, si y sólo si los estados mentales del B posterior (esto es, los recuerdos de B, sus creencias, carácter, proyectos, intenciones, deseos, etc.) son descendientes causales de los estados mentales de A. Así, B podría recordar una experiencia tenida por A, B podría actuar para cumplir una aspiración de A, el sentido del humor de B podría haber sido heredado de A, y así sucesivamente. Si existe un número suficientemente grande de vínculos psicológicos, o de estratos superpuestos de tales vínculos, podemos hablar de continuidad psicológica. La continuidad psicológica es una relación transitiva, pero las conexiones psicológicas mismas (es decir, los lazos psicológicos directos que forman la continuidad) son intransitivas[9]. De ahí que yo pueda ser psicológicamente continuo con alguien psicológicamente muy diferente de mí.

Mientras que los criterios corporal y cerebral son análisis de la identidad personal obviamente reductivos, es discutible si el criterio psicológico es reductivo o no. Ello depende de si la noción de continuidad psicológica puede o no ser enteramente comprendida sin hacer referencia a personas. ¿Se pueden describir los dolores y los pensamientos sin hacer referencia a los sujetos que los tienen? ¿No es esencial a los contenidos de recuerdos e intenciones la refe-

rencia a personas (yo recuerdo que *yo* me comí un helado, yo quiero que *yo* salga de vacaciones el mes que viene, etc.)? Estas preguntas figuran entre las más profundas del tema de la identidad personal.

Sin embargo, podemos plantearnos antes la pregunta de si alguno de estos criterios es plausible. Los criterios, recuérdese, eran formulados para revelarnos la esencia de la identidad personal. Son respuestas a la pregunta constitutiva, e intentan ser verdades necesarias, verdaderas respecto de cualquier persona en cualquier situación posible. Así pues, si existiese un mundo posible, no importa cuán diferente del nuestro, en el que un lado del criterio fuera verdadero y el otro falso, el criterio sería falso.

He aquí un experimento mental que describe un escenario lógicamente posible y que, en opinión de muchos, refuta el criterio corporal. Sydney Shoemaker fue el primero en introducir en la bibliografía sobre el tema este experimento mental (al que denominó «El transplante cerebral»). Escribió:

> Ahora [que] es posible transplantar ciertos órganos [...] es al menos concebible [...] que un cuerpo humano pueda continuar funcionando normalmente si su cerebro es reemplazado por otro tomado de otro cuerpo humano. [...] Dos hombres, llamados el señor Brown y el señor Robinson, son operados de tumores cerebrales, y a ambos se les practica una extracción de cerebro. Al final de las operaciones, sin embargo, el asistente coloca sin darse cuenta el cerebro de Brown en la cabeza de Robinson, y el cerebro de Robinson en la cabeza de Brown. Uno de estos hombres muere instantáneamente, pero el otro, que tiene la cabeza de Robinson y el cerebro de Brown, recupera después de un tiempo la consciencia. Denominemos a éste «Brownson» [...] Cada vez que se le pregunta cómo se llama, responde «Brown». Reconoce a la esposa y a la familia de Brown [...] Y es capaz de describir en detalle sucesos de la vida de Brown [...] pero no da muestras de conocer nada de la vida de Robinson[10].

Casi todo el mundo está de acuerdo en que la mejor manera de describir esta situación es afirmando que Brown es Brownson. Pero algunos piensan: Brown muere y Robinson adquiere un cerebro y una psicología nuevos. Así pues, es razonable reconocer que el caso del «trasplante de cerebro» refuta el criterio corporal, y que, por consiguiente, refuta el animalismo (la concepción de que somos idénticos a seres humanos), ya que Brown no es el mismo ser humano que Brownson y, sin embargo, es la misma persona.

¿Qué otra moraleja deberíamos inferir de esta refutación del criterio corporal? Para empezar, que tiene incidencia en las otras dos teorías disponibles. Quizás Brown sea Brownson porque Brownson tiene el cerebro de Brown. O quizás Brown sea Brownson porque Brownson es psicológicamente continuo con Brown. Pero hay otro experimento mental que suscita problemas tanto para el criterio cerebral como para el criterio psicológico: el de la «fisión».

Imaginemos que mi cerebro fuera dividido a la mitad, y que cada uno de los dos hemisferios (equipolentes) fuera transplantado, respectivamente, a dos

cuerpos (sin cerebros)[11]. Tras la operación, despiertan dos individuos —a los que denominaremos «Ízquier» y «Dérech»—, cada uno de los cuales es psicológicamente continuo conmigo (ambos tienen mis creencias y carácter, y ambos parecen recordar mi vida pasada). En el momento en que despiertan, son exactamente iguales entre sí desde el punto de vista psicológico, pero a partir de ahí comienzan a diferir.

Entiendo la «fisión» como un escenario lógicamente posible; es una variante del «transplante cerebral». ¿De qué forma representa la fisión un problema para el criterio cerebral y el criterio psicológico? El criterio psicológico implica que yo soy Ízquier y que yo soy Dérech. (ya que soy psicológicamente continuo con ambos). En consecuencia, de acuerdo con la transitividad de la identidad, se sigue que Ízquier es Dérech. Pero eso no puede ser correcto[12]. Ízquier y Dérech, aunque inicialmente hayan sido muy similares entre sí, son dos personas y no una.

Quizás se piense que para el criterio cerebral no se presenta un problema análogo. Este criterio no implicaría que yo sea Ízquier o Dérech, ya que ni Ízquier ni Dérech están en posesión de todo mi cerebro. Pero el criterio cerebral, para ser plausible, no puede requerir la identidad del cerebro entero como condición de la identidad personal. Hay personas que han sobrevivido a pesar de que fue destruida la mitad de su cerebro. Si el hemisferio restante de una persona así fuera transplantado a un nuevo cuerpo, consideraríamos ese nuevo estado como una continuación de la vida de la persona. Por lo tanto, para que tenga verosimilitud, el criterio cerebral debe ser entendido en los términos siguientes:

Criterio cerebral revisado: A es en $t_1$ la misma persona que B en $t_2$, si, y sólo si, B tiene una cantidad suficiente del cerebro de A para sostener la continuidad psicológica.

La «fisión» es un problema para el criterio cerebral revisado. Ese criterio, al igual que el criterio psicológico, implica que soy Ízquier y que soy Dérech (ya que ambos poseen la cantidad suficiente de mi cerebro para sostener la plena continuidad psicológica).

¿Deberíamos concluir entonces que el criterio cerebral revisado y el criterio psicológico son falsos? Eso sería demasiado apresurado. Se ha gastado mucha tinta defendiendo estos criterios contra la presente objeción. Una vía —la respuesta de ocupación múltiple— opta por conservar los criterios en su estado actual, pero alegando que el caso de la fisión ha sido inadecuadamente descrito. La otra vía —la respuesta de la singularidad— implica modificar los criterios insertando en los análisis una cláusula de no ramificación. Ninguna de estas respuestas es cónsona con el sentido común.

## La fisión

El escenario descrito a continuación es lógica, aunque aún no técnicamente posible. Unos cirujanos dividen mi cerebro y transplantan sus hemisferios en sendos cuerpos carentes de cerebros. Dos individuos se despiertan —llamémoslos Dérech e Ízquier— que son iguales a mí desde el punto de vista psicológico. Ambos tienen el mismo carácter, el mismo sentido del humor, las mismas convicciones políticas conservadoras, etc., y ambos tienen aparentemente recuerdos de mi pasado. Algunos filósofos piensan que los casos de fisión nos revelan algo acerca de la naturaleza de la identidad personal y acerca de su importancia. En primer lugar, se afirma que, dado que yo no soy ni Dérech ni Ízquier, pero continuaría existiendo aunque sólo uno de éstos hubiera sobrevivido, se sigue que la identidad personal consiste en una cierta continuidad (física o psicológica) de tipo biunívoco o no ramificada. En segundo lugar, se afirma que mi relación hacia Ízquier y Dérech contiene todo lo que es relevante en casos normales de identidad. Puesto que yo no soy idéntico a Ízquier o a Dérech, la identidad no puede ser lo determinante. Lo que importa es tener un continuador psicológico, independientemente de si este continuador es uno mismo o no. Parfit acepta las dos implicaciones mencionadas. Otros aceptan la primera, pero no la segunda. Los defensores de la llamada concepción simple rechazan ambas implicaciones.

## Defensa de la concepción compleja

Comencemos con la respuesta de la ocupación múltiple, defendida por Lewis, Noonan y otros[13]. De acuerdo con esta respuesta, mi fisión ha sido mal descrita. No involucra a tres personas (yo, Ízquier y Dérech), sino sólo a dos (Ízquier y Dérech). Ízquier y Dérech existen ambos ya antes de la fisión, ocupando el mismo cuerpo («mi» cuerpo). Tras la fisión, cada uno pasa a ocupar su propio cuerpo.

Sin embargo, no está claro cómo esta redescripción de la fisión ayuda a conservar el criterio psicológico (como se proponen hacer Lewis y Noonan). Si tanto Ízquier como Dérech existen con anterioridad a la fisión, ¿no es acaso cierto que Ízquier antes de la fisión es psicológicamente continuo con Dérech después de la fisión, y que Dérech antes de la fisión es psicológicamente continuo con Ízquier después de la fisión? Pero en ese caso, el criterio psicológico implicaría que Ízquier es Dérech. Ahora bien, habíamos redescrito «fisión» precisamente para evitar esta consecuencia. ¿Qué hemos, pues, ganado?

En segundo lugar, casi huelga decir que la explicación de la ocupación múltiple está metafísicamente viciada. Se nos pide que creamos que dos personas ocupan un único cuerpo, a pesar de que todos los criterios usuales indican que sólo una persona está presente. En tercer lugar, «mis» oraciones de «yo» previas a la fisión serían ambiguas, pues tendrían como referente tanto a

Ízquier como a Dérech. ¿Cómo se explicaría entonces la aparente unidad de los pensamientos en primera persona asociados con el cuerpo previo a la fisión? (No parece que mis pensamientos en primera persona previos a la fisión fueran ambiguos.) ¿Cómo determinar si algún uso del pronombre «yo» se estaba refiriendo a Ízquier o, más bien, a Dérech?

Queda por evaluar la respuesta de la singularidad. Esta respuesta ha sido suscrita por Parfit (en defensa del criterio psicológico). Éste escribe:

Williams ha atacado [el criterio psicológico] con el argumento siguiente. La identidad es una relación biunívoca [*one-one relation*]. Por lo tanto, cualquier criterio de identidad debe apelar a una relación que sea lógicamente biunívoca. La continuidad psicológica no es biunívoca. Por lo tanto, no puede servir de criterio. [Esto es una reformulación de la presente objeción.] [...] Algunos escritores han respondido que es suficiente si la relación utilizada siempre es, de hecho, biunívoca [...] Sugiero una réplica ligeramente diferente. La continuidad psicológica es una razón para hablar de identidad cuando es biunívoca[14].

En otras palabras, se nos está proponiendo que analicemos la identidad personal a lo largo del tiempo en términos de una continuidad psicológica **no ramificada**, una relación que, por definición, sea lógicamente biunívoca. Por lo tanto, deberíamos suscribir también lo siguiente:

Criterio psicológico revisado: A en $t_1$ es la misma persona que B en $t_2$, si y sólo si A y B son psicológicamente continuos y el flujo de la vida mental no se ha ramificado entre $t_1$ y $t_2$.

De acuerdo con este criterio revisado, no sería idéntico ni a Ízquier ni a Dérech, puesto que ha ocurrido una ramificación en el caso de la fisión. Por lo tanto, eludimos el resultado poco intuitivo de que Ízquier es Dérech.

Sin embargo, el criterio psicológico revisado tiene una consecuencia contraintuitiva. (Daré por sentado que una versión modificada del criterio cerebral revisado —esto es, que A en $t_1$ es la misma persona que B en $t_2$, si, y sólo si, B tiene una porción suficiente del cerebro de A para mantener la continuidad psicológica —también tiene esta consecuencia[15].)

Según el criterio psicológico revisado, no soy ni Ízquier ni Dérech, pero si una enfermera hubiera dejado caer el hemisferio izquierdo durante la operación haciendo que se destruyera, y haciendo así que hubiera un solo superviviente, entonces yo habría sobrevivido. Si yo no fuera ni Ízquier ni Dérech en el mundo de la fisión, pero sobreviviese en el mundo de un sólo superviviente, parece que tengo una razón poderosa para sobornar a la enfermera de modo que ésta deje caer al suelo uno de los hemisferios, asegurando así mi supervivencia. Pero si, por ejemplo, tengo una buena razón para pensar que la operación del hemisferio derecho será exitosa, ¿debería realmente preocuparme la posibilidad de que otra operación, que ocurriese, digamos, en otro edificio, tuviera éxito? ¿En qué podría afectarme eso? Pero ocurre que sí me afecta:

para mí sería asunto de vida o muerte. Así pues, según el criterio psicológico revisado, algo que no debería afectarme de hecho me afecta.

No obstante, Parfit dispone de una réplica ingeniosa contra esta objeción. Ya en un trabajo más temprano (del año 1971), no se había sentido a gusto con la descripción de la fisión como algo que implica que no soy idéntico ni a Ízquier ni a Dérech. Inicialmente creyó que no había descripción adecuada alguna de la fisión en términos de identidad, y que, de producirse la fisión, «nos veríamos forzados [...] a abandonar el discurso de la identidad»[16]. Oponiéndose a la descripción de que no soy ni Ízquier ni Dérech, arguía de la forma siguiente:

> Estuvimos de acuerdo en que yo podía sobrevivir si mi cerebro fuera transplantado exitosamente. Y hay gente que de hecho ha sobrevivido a la destrucción de la mitad de su cerebro. Parece seguirse que yo podría sobrevivir si la mitad de mi cerebro fuese transplantada exitosamente y la otra mitad fuese destruida. Pero si esto es así, ¿por qué yo *no* habría de poder sobrevivir si la otra mitad fuese transplantada exitosamente? ¿Cómo podría un doble éxito constituir un fracaso?[17].

Para el momento de publicar la versión en inglés de su libro *Razones y personas* (1984), Parfit ya no estaba convencido de este argumento, y reconoció que «yo no soy ni Ízquier ni Dérech» era la mejor descripción posible de la fisión. Y tenía razón, pues el argumento de 1971 no es sólido. Saca partido de una ambigüedad de la palabra «éxito». En un sentido puramente médico, un procedimiento es exitoso si se desarrolla según lo planeado. En ese sentido, es obvio que un doble éxito no es un fracaso. Pero si el «éxito» requiere la supervivencia de la persona inicial, entonces un «doble éxito» puede representar, de hecho, un fracaso. Un éxito en el sentido médico no necesariamente supone un éxito en el sentido de la supervivencia personal.

Sin embargo, hay otra reflexión que está presente tanto en el artículo de 1971 como en el libro de 1984, y que es la siguiente: es absurdo asimilar la perspectiva de una fisión al de la muerte ordinaria. A menos que saque de quicio la idea de la fisión, limitarse a afirmar que no soy ni Ízquier ni Dérech sería engañoso. Parfit piensa que una fisión equivale a la supervivencia ordinaria. Mi relación con Ízquier, y mi relación con Dérech, contienen todos los atributos de la supervivencia ordinaria. Puesto que no soy idéntico ni a Ízquier ni a Dérech (como hemos concedido ahora), se sigue que lo que importa no es la identidad. Lo que importa es la continuidad y/o el estar vinculado.

La tesis de que la identidad no es lo que realmente importa provee a Parfit de una réplica a nuestra objeción al criterio psicológico revisado. Si la identidad no importa, no tengo razón para sobornar a la enfermera para que deje caer uno de los hemisferios. No tengo razón alguna para dar origen a un mundo en el que sobreviva un solo individuo en lugar de un mundo en el que sobrevivan dos, incluso si yo pudiese sobrevivir (continuar existiendo) únicamente en el primer mundo. Con tal de que hubiera al menos un superviviente

con el que yo tuviese una continuidad psicológica completa, no importaría si hay más de uno.

## ¿Es importante la identidad?

Es obvio, pues, que Parfit necesita la tesis de que la identidad no es lo decisivo, si quiere refutar la objeción al criterio psicológico revisado. Pero ¿en qué consiste exactamente esta tesis? En realidad, son dos las tesis por considerar: la tesis negativa de que la identidad no es lo importante y la tesis positiva de que la relación de continuidad psicológica y/o conectividad psicológica (lo que Parfit denomina la relación R) es lo que importa[18].

Supongamos que yo estoy a punto de dividirme mañana, y que sé que Dérech tendrá dolor de muelas. Parfit piensa que, dada la relacionalidad-R [*R-relatedness*] que Dérech tiene conmigo, yo debería preocuparme por el dolor de muelas que Dérech va a sufrir mañana como si fuera yo quien fuera a sufrir el dolor. En cuanto a la identidad, sería algo irrelevante.

Pero ¿en qué consistiría mi preocupación por Dérech y por mí? Cuando me intereso por mi futuro, me estoy interesando en mí mismo, y es evidente que cuando me intereso por el futuro de Dérech no me estoy interesando en mí mismo, ya que él no es yo. Así pues, parece que yo no puedo tener el mismo tipo de interés por mí que el que tengo por Dérech. Parfit respondería a esto que se trata de un punto meramente verbal. Es verdad que no puedo decir que mi interés por Dérech sea un interés en mí mismo, pero se trataría esencialmente del mismo tipo de interés que el que tengo hacia mí mismo.

Esto presupondría que el *interés en uno mismo* es un concepto compuesto, integrado por los elementos (supuestamente) más básicos de la *identidad* y el *interés,* en que este último no presupondría la identidad, y equivaldría al tipo de interés que yo puedo tener hacia cualquier individuo con quien me halle en la relación R. En consecuencia, puedo considerar el interés que tengo hacia mí mismo y el interés que tengo hacia Dérech como esencialmente iguales, salvo por una diferencia en su descripción verbal.

La cuestión crucial, empero, y hasta ahora irresuelta, es la de si el *interés en uno mismo* es un concepto unitario, no reducible a conceptos más elementales, o si es un concepto compuesto, integrado precisamente por conceptos de este tipo. Sólo si el *interés en uno mismo* resultara ser un concepto compuesto, podría funcionar la defensa que Parfit hace del criterio psicológico revisado.

La sustentabilidad del reduccionismo de Parfit depende de consideraciones del mismo tipo. Según el reduccionismo, la vida psicológica de una persona puede ser descrita completamente en términos impersonales (es decir, sin hacer referencia a personas o a la identidad personal). Esto puede parecer ya poco plausible con respecto a simples estados mentales, como sufrimientos y dolores de muelas, que por su propia naturaleza parecen exigir un propietario

o un portador. Pero es que además hay un problema específico con respecto a los estados psicológicos tales como los recuerdos y las intenciones, que parecen llevar incorporada la identidad en su contenido. Así, mi recuerdo de una experiencia pasada (digamos, comerse un helado ayer) no sólo requiere un portador presente, sino que parece incluirme a mí en su contenido: yo recuerdo que *yo* me comí un helado ayer. (Por eso, si resultase que yo en realidad nunca me comí un helado, tú podrías perfectamente responder «tú no recuerdas haber comido helado, simplemente crees que lo recuerdas».) ¿Cómo, entonces, se podría describir el recuerdo, uno de los rasgos fundamentales de nuestra vida psicológica, de forma impersonal?

La respuesta de Parfit es que la memoria es un concepto compuesto construido a partir de la *identidad* y de un *cuasirecuerdo (recuerdo-q)*. Este último recuerdo es definido como algo similar a un recuerdo en todos los aspectos fenoménicos y causales, salvo por el hecho de que no presupone la identidad. Por consiguiente, puedo tener recuerdos-q de las experiencias de otro individuo, y lo que denominamos recuerdos son sólo recuerdos-q de nuestras propias experiencias. Al igual que lo que sucedía con el *interés en uno mismo*, la cuestión es si el recuerdo es efectivamente un concepto compuesto, o si es más bien un concepto unitario no reducible a átomos conceptuales más básicos. (Una duda, por ejemplo, es si el *recuerdo-q* puede realmente ser entendido independientemente del *recuerdo*. De no ser así, el recuerdo implicaría, después de todo, la identidad.) Este debate, al igual que el referido al *interés en uno mismo*, sigue abierto.

## La concepción sencilla

¿Qué conclusión deberíamos extraer de nuestra discusión anterior? Sería apresurado concluir que la concepción compleja es insostenible. A pesar de ello, ha sido atacada duramente, y se encuentra a la defensiva. ¿Hay consideraciones ulteriores que hablen en su contra y que, por lo tanto, nos inclinen hacia la concepción sencilla?

Según la concepción sencilla, la relación de la identidad personal es primitiva e inanalizable. Butler sostuvo que la palabra «igual» es usada en un «sentido estricto y filosófico» cuando es aplicada a personas, pero en un sentido «laxo y popular» cuando es aplicada a los cuerpos y a otro tipo de cosas (por ejemplo, a los artefactos). De forma similar, Reid escribió sobre la identidad lo siguiente:

> no tiene una naturaleza fija cuando se aplica a los cuerpos, y muchas veces las preguntas sobre ella son preguntas verbales. Pero cuando la identidad se aplica a las personas, no es ambigua ni admite grados del más y el menos. La identidad es el fundamento de todos los derechos y las obligaciones, así como de la responsabilidad, y su noción es fija y precisa[19].

Así pues, mientras que llamar o no a este coche «el mismo coche» que algún coche previo podría ser una cuestión verbal, es decir, un asunto solucionable por convención, nunca sería una cuestión verbal la de si una persona en un tiempo determinado es la misma que otra persona en otro tiempo determinado, y tal cuestión jamás podría ser zanjada por convención.

En su artículo de 1971, Parfit comenzaba afirmando que hay cierta opinión sobre la naturaleza de la identidad personal, y otra sobre su importancia[20]. Ambas opiniones, decía, son falsas. La opinión sobre la naturaleza de la identidad personal es que siempre se trata de una cuestión de todo o nada. O yo existiré en una situación futura o no lo haré. No hay una zona gris. Parfit señalaba que no pensamos que esto sea así en el caso de las naciones o de las máquinas. No pensamos que tiene que haber una **respuesta «de sí o no»** a la pregunta «¿siguió siendo Inglaterra la misma nación después de 1066?». Tampoco tendría que haber, añadía, una respuesta de ese tipo en el caso de las personas[21].

Todas las versiones de la concepción compleja implican que puede haber casos de indeterminación en la identidad personal (es decir, casos en los que no hay un respuesta «de sí o no» sencilla)[22]. Por ejemplo, considérese el criterio psicológico revisado. Si yo experimento algunos cambios psicológicos menores mañana, soy psicológicamente continuo con la persona que ocupa mi cuerpo mañana. Esto es un caso claro de identidad, de acuerdo con el criterio psicológico revisado. Si, por otra parte, yo experimento cambios psicológicos masivos, que incluyen un cambio radical de carácter, implantación de nuevos «recuerdos», etc., entonces no soy psicológicamente continuo con la persona que ocupará mi cuerpo mañana, y, por lo tanto, es indudable que no soy idéntico a esa persona. Pero hay casos intermedios, en los que no está definido si yo soy psicológicamente continuo con la persona de mañana. En tales casos, el criterio psicológico revisado no permite determinar si yo soy idéntico o no a la persona posterior. La pregunta «¿soy yo esa persona?» no tiene una respuesta «de sí o no».

Parfit, por supuesto, no tendría empacho en admitir esta consecuencia del criterio psicológico revisado, ya que ayuda a subvertir la creencia común sobre la naturaleza de la identidad personal, lo cual era uno de sus objetivos. Pero aunque es fácil entender la indeterminación de la identidad en el caso de barcos y coches, es difícil entenderla en el caso de las personas, especialmente cuando consideramos la cuestión desde el punto de vista de la primera persona. Bernard Williams fue el primero en insistir en este punto en su esclarecedor artículo «El yo y el futuro»[23].

Williams señalaba que no disponemos de un modelo para prever la indeterminación, o reaccionar emocionalmente ante ella, cuando se trata de nosotros mismos:

> El que se me diga que una situación futura representa un caso límite respecto de si seré yo quien vaya a ser herido —o sea, que es conceptualmente indeterminable si se tratará o no de mi persona— es algo que parece dejarme atónito; en especial, porque mis expectativas y las emociones que las acompañan no parecen dar cabida a un concepto semejante[24].

Un caso indeterminado o conceptualmente indeterminable no es uno en el que, por ejemplo, se me diga que uno de los dos individuos que ocupamos esta habitación recibirá un tiro mañana. En ese caso, yo conozco las dos posibilidades que existen: o me pegarán un tiro o no. Ni tampoco se parece al caso en el que yo pienso que un «horror indescriptible» me afectará: independientemente de cuál sea el horror, éste me afectará o no. Pero si se me dice que mañana alguien sufrirá un gran dolor y, luego, que no es determinable si esa persona seré yo, no tengo idea de cómo reaccionar (¿con miedo?, ¿compasión?, ¿angustia?). No hay una reacción cognitiva o emocional propia de esta situación. «Una obstinada frustración sigue al empeño de articular en mis expectativas una situación en la que es conceptualmente indeterminable si yo estaré presente o no.»[25]

Este problema para la concepción compleja no debería ser subestimado. Puede que no refute dicha concepción, pero hace que algunos de los juicios que se derivan de ella sean difíciles de entender. Por contraste, la concepción sencilla no presenta problemas de ese tipo. Según ésta, la identidad personal es algo de todo o nada, que nunca admite grados. A diferencia de la identidad de otras cosas, nunca podría haber situaciones en las que una sombra conceptual se extendiese sobre nuestra identidad[26].

Adicionalmente, la concepción sencilla no presupone ni la ocupación múltiple ni la inclusión de una cláusula de no ramificación o de exclusión de competidores (y, por lo tanto, no tiene que coquetear con la idea de que «la identidad no es lo que importa»). Para la concepción sencilla, la «fisión», como la describimos arriba, deja abierta la cuestión de la identidad. Yo podría ser Ízquer, o también Dérech, o podría no ser ninguno de los dos. Cualquier descripción en términos de continuidades físicas o psicológicas fallaría definitivamente en cuanto a zanjar la pregunta de quién es quién. La identidad persona sería siempre un «hecho que va más allá de» tales continuidades. Cualquiera que estuviera de acuerdo en someterse a una fisión estaría corriendo un riesgo, pero, como Williams observa respecto de otro caso enigmático, «el hecho de que haya cabida aquí para la noción de riesgo es en sí mismo un aspecto importante del problema»[27].

A veces se critica la concepción sencilla diciendo que hace que la identidad personal a través del tiempo se vuelva incomprensible. No hay tal cosa. Es verdad que, según esa concepción, nuestra evidencia ordinaria de que hay identidad personal a través del tiempo (la presencia de varias continuidades) no sería garantía lógica de la identidad. Pero todos damos por sentado que poseemos conocimiento del mundo exterior y de otras mentes a pesar de que nuestra evidencia no garantiza lógicamente la existencia del mundo exterior y de otras mentes. ¿Por qué entonces el conocimiento de nuestra identidad y de la de los demás debería ser amenazada por la impugnabilidad lógica [*logical defeasibility*] de nuestra evidencia?

Asumiendo que la concepción simple sea verdadera, ¿qué es para ella una persona? Una persona no podrá ser idéntica a su cerebro, a su cuerpo o a cual-

quier otra entidad biológica (pues entonces alguna variante de la concepción compleja sería verdadera). Para la concepción sencilla, la existencia de una persona quizás dependa causalmente de la existencia y el normal funcionamiento de su cerebro, pero no necesita darse una dependencia metafísica. Tal concepción es ciertamente compatible con la tesis de que las personas son almas inmateriales, pero no es obvio que la requiera. Finalmente, la concepción sencilla cuadra bien con la tesis de que somos esencial y fundamentalmente personas.

## Observaciones finales

Hemos cubierto una gran superficie de terreno en este capítulo. Pudimos distinguir los sentidos de «identidad» numéricos de los cualitativos, y las cuestiones constitutivas de las evidenciales. Examinamos la cuestión de la naturaleza de una persona, y luego diversas explicaciones sustanciales de lo que significa que una persona persista en el tiempo, en todas las cuales encontramos deficiencias. El criterio psicológico revisado quizás sea la versión más plausible de la concepción compleja, que mejor refleja la concepción de las personas como seres mentales; pero su defensa requiere la verdad de la tesis polémica de que la identidad no es lo que importa. Es más, la posibilidad de la indeterminación en la identidad personal, que es un rasgo común de todas las variantes de la concepción compleja, es profundamente problemática. La concepción sencilla tiene la virtud de evitar estas dificultades y, en general, es una concepción de las personas que merece más atención de la que ha recibido en las discusiones contemporáneas.

## Preguntas para el estudio

- ¿Es una persona fundamentalmente un ser con actividad mental?
- ¿Refuta el ejemplo del «transplante cerebral» el criterio corporal?
- ¿Son aceptables las consecuencias del criterio psicológico revisado?
- ¿En qué consiste la tesis de que la identidad no es importante?
- ¿Cuál es el precio de aceptar la concepción sencilla?

### Notas

1. Véanse J. Butler, «Of Personal Identity», reimpreso en J. Perry (ed.), *Personal Identity* (Los Ángeles, Calif.: University of California Press) 1975; T. Reid, *Essays on the Intellectual Powers of Man,* ed. de A. Woozley (Londres: Macmillan) 1941, y R. Swinburne, «Personal Identity: The Dualist Theory», en S. Shoemaker y R. Swinburne, *Personal Identity* (Oxford: Basil Blackwell) 1984.
2. Véanse D. Parfit, «Personal Identity», reimpreso en J. Glover (ed.), *The Philosophy of Mind* (Oxford: Oxford University Press) 1976, pp. 142-163, y *Reasons and Persons* (Oxford: Oxford University Press) Parte III.

3. J. Locke, *Essay Concerning Human Understanding,* ed. W. Carroll (Bristol: Thoemmes) 1990, II, xxvii, 9.
4. Véase, por ejemplo, R. Descartes, *Meditations on First Philosophy,* ed. J. Cottingham (Cambridge: Cambridge University Press) 1996.
5. David Hume, *A Treatise of Human Nature* (1739), Parte IV, Sección IV: Of Personal Identity. Reimpreso en A. J. Ayer y R. Winch (eds.), *British Empirical Philosophers* (Londres: Routledge & Kegan Paul) 1965, p. 490.
6. Dicho a grandes rasgos, F es un sortal sustancial si nos dice lo que algo es fundamental o esencialmente. Si *x* cae bajo el sortal sustancial F, entonces *x* es fundamentalmente F. Por contraste, G es un sortal de fase si *x* cae o podría caer bajo G durante una parte de su existencia, aunque no durante toda ella. Así, «maestro», «filósofo» y «monociclista» son ejemplos inequívocos de sortales de fase. Ejemplos relativamente inequívocos de sortales sustanciales son, en cambio, «escritorio», «coche», «barco», «árbol», «perro», etc. Véase D. Wiggins, *Sameness and Substance* (Oxford: Basil Blackwell) 1980, pp. 62-66.
7. Para una defensa de esta concepción, véanse P. Snowdon, «Persons, Animals and Ourselves», en *The Person and the Human Mind,* ed. de C. Gill (Oxford: Clarendon Press) 1990, y E. Olson, *The Human Animal* (Oxford: Oxford University Press) 1997.
8. Es decir, por ejemplo, X en $t_1$ es el mismo cerebro que Y en $t_2$, si, y sólo si, un trayecto continuo puede ser reconstruido desde X en $t_1$ hasta Y en $t_2$.
9. O sea, si A es psicológicamente continuo con B, y B es psicológicamente continuo con C, entonces A es psicológicamente continuo con C. Pero A podría estar muy fuertemente conectado de forma psicológica con B, y B con C, sin que A tenga que estar muy fuertemente conectado de forma psicológica con C (por ejemplo, en el caso de los recuerdos, los cuales se desdibujan con el tiempo).
10. S. Shoemaker, *Self-Knowledge and Self-Identity* (Ithaca, NY: Cornell University Press) 1963, pp. 23-24.
11. La mayoría de nosotros no tiene hemisferios equipolentes o funcionalmente equivalentes; pero habríamos podido tenerlos. Se supone que los criterios cerebrales y psicológicos han de poder ser aplicados a todas las personas, incluyendo aquellas que poseen hemisferios equipolentes.
12. La transitividad de la identidad estipula que si X = Y e Y = Z, entonces X = Z.
13. Véanse D. Lewis, «Survival and Identity», reimpreso en *The Identities of Persons,* ed. de A. O. Rorty (Los Ángeles, Calif.: University of California Press) 1969, pp. 17-41, y H. Noonan, *Personal Identity* (Londres: Routledge) 2003, pp. 139-143.
14. Parfit, «Personal Identity», *op. cit.*, p. 149.
15. Hay otras objeciones al criterio cerebral (en todas sus variantes). Tal como fue formulado arriba, el concepto de una persona es el concepto de una clase bastante sofisticada de ser con actividad mental. Por lo tanto, no tiene muchas perspectivas de éxito cualquier teoría que entienda la identidad personal a través del tiempo en términos de la existencia continuada de una entidad no esencialmente dotada de mente (como un cuerpo, un cerebro o un ser humano). En segundo lugar, la única razón por la que el cerebro es seleccionado como un elemento esencial es porque permite la continuidad psicológica. En ese caso, su importancia sería totalmente derivativa y el cerebro no podría constituir la esencia de la identidad personal.
16. Parfit, «Personal Identity», *op. cit.*, p. 149.
17. Ibíd., p. 144.

18. Parfit piensa que esta combinación de concepciones tiene harta relevancia para la moralidad y la racionalidad. Por ejemplo, considera que si la identidad no es lo que importa, entonces el hecho de la «separación de las personas» tiene menor trascendencia que de costumbre. Esto, a su vez, implicaría que se debe atribuir menos peso a los principios distributivos. Debemos aspirar a maximizar la suma neta de beneficios una vez restadas las cargas, sin importar cuál sea su distribución. Asimismo, si la conectividad psicológica es lo que importa, una persona puede merecer menos castigo cuanto menos psicológicamente conectado esté con su yo criminal anterior (por ejemplo, si se ha reformado realmente). En el caso de la racionalidad, la menor importancia de la identidad implicaría que el principio de interés en sí mismo (de acuerdo con el cual es «especialmente racional actuar en beneficio propio») «carece de fuerza» («Personal Identity», *op. cit.*, p. 161).
19. T. Reid, *Essays on the Intellectual Powers of Man,* ed. de A. Woozley (Londres: Macmillan) 1941, citado en Noonan, *op. cit.*, p. 16.
20. Véase también su discusión con Godfrey Vesey, «Brain Transplants and Personal Identity», en *Philosophy in the Open* (Milton Keynes: Open University Press) 1978.
21. Parfit, «Personal Identity», *op. cit.*, p. 146. Su ejemplo, en el artículo de 1971, de un caso en el que la pregunta de la identidad personal no recibe una respuesta «de sí o no» era el de la fisión. Pero, como vimos, éste podría no ser el mejor ejemplo para ese fin. De hecho, la fisión desempeña un papel mucho más importante en el socavamiento de la creencia sobre la importancia de la identidad personal.
22. Nótese que la indeterminación reside en el valor de verdad de las oraciones o los enunciados de identidad personal. Doy por sentado que Gareth Evans («Can There Be Vague Objects?», *Analysis,* vol. 38, 1978, p. 208) ha demostrado que la relación de identidad, por sí sola, nunca es una fuente de vaguedad o indeterminación. Cualquier vaguedad de una oración de identidad se debe a una vaguedad en uno o ambos términos singulares (es decir, lo vago es a cuál objeto selecciona el término respectivo).
23. B. Williams, «The Self and the Future», en su colección *Problems of the Self* (Cambridge: Cambridge University Press) 1982, pp. 46-64, especialmente pp. 58-63.
24. Ibíd., p. 58.
25. Ibíd., p. 61.
26. La frase «sombra conceptual» se debe a Williams, Ibíd., p. 60.
27. Ibíd., p. 63.

## Bibliografía adicional comentada

H. Nooman, *Personal Identity,* segunda edición (Londres: Routledge) 2003. Una óptima introducción, aunque exigente, a la historia del tema y al debate contemporáneo sobre él.

D. Parfit, *Reasons and Persons* (Oxford: Oxford University Press) 1984, Parte III. Este libro ha dictado en buena parte la agenda de las discusiones recientes sobre la identidad personal. Aunque su estilo no es introductorio, Parfit escribe de manera clara, y cualquier lector interesado debería ser capaz de seguir el curso de sus ideas radicales y revisionistas.

S. Shoemaker y R. Swinburne, *Personal Identity* (Oxford: Basil Blackwell) 1984. Un debate animado entre uno de los principales defensores de la concepción compleja y un partidario reconocido de la concepción sencilla.

## Recursos en Internet

B. Garrett (1998, 2004) «Personal Identity», en E. Craig (ed.), *Routledge Encyclopedia of Philosophy.* Bajado el 31 de mayo de 2006 de <http://www.rep.routledge.com/article/V024>.

E. T. Olson (2002) «Personal Identity», *The Stanford Encyclopedia of Philosophy (edición de otoño de 2002),* Edward N. Zalta (ed.). Bajado el 31 de mayo de 2006 de <http://plato.stanford.edu/archives/fall2002/entries/identity-personal>.

# 9. El realismo y el antirrealismo

## Introducción

Por realismo sobre algún tema, como los F, se entiende normalmente la concepción de que los F existen, y que lo hacen independientemente de nosotros. La «independencia» no es una noción transparente, pero podemos afirmar que los F existen independientemente de nosotros si se cumple alguna de las tres condiciones siguientes:

(i) Aunque nosotros no hubiésemos existido, los F existirían.
(i) Aunque dejáramos de existir, los F seguirían existiendo.
(iii) La naturaleza de los F no está determinada únicamente por la manera en que nosotros concebimos que ésta es[1].

Intuitivamente, somos realistas [*realists*] acerca de los planetas ([i]-[iii] son todos enunciados verdaderos acerca de los planetas), pero irrealistas [*non-realists*] acerca de la moda (ni [i], ni [ii] ni [iii] es verdadero sobre la moda: algo está de moda sólo porque la gente tiene determinada actitud hacia ello).

La oposición hacia el realismo de los F puede, por consiguiente, adoptar dos variantes: la insistencia en que los F no existen o la insistencia de que, aunque algunos F existen, no lo hacen independientemente de nosotros. Las teorías correctivas [*error theories*] (como las que han sido formuladas sobre el discurso moral y la aritmética) y el expresivismo (formulado sobre el dis-

curso moral) son ejemplos del primer tipo de estrategia irrealista[2]. El idealismo es un ejemplo del segundo tipo de estrategia. El obispo Berkeley pensaba que los planetas existían, pero que, como estaban compuestos de ideas, no existían independientemente de las mentes (mentes de los hombres o de Dios). En este capítulo me propongo examinar un ejemplo más reciente del segundo tipo de estrategia irrealista, introducido por primera vez en la década de 1960 por Michael Dummett. Dummett denomina «antirrealismo» [*anti-realism*] a su clase particular de irrealismo, con el cual se puede decir que trata de forjar un nuevo sentido en que la existencia de los F podría no ser independiente de nosotros.

Dummett sugiere que las disputas tradiciones de la metafísica podrían recibir nuevas luces, o incluso soluciones, si se desarrollasen dentro del marco del debate entre el realismo y el antirrealismo. Aunque entendió a este último como un debate sobre el significado y la verdad, creía que el mismo podía aclarar las disputas metafísicas tradicionales respecto de la naturaleza de las matemáticas, las otras mentes, los objetos materiales y la realidad del pasado y del futuro.

Basándome en dos artículos tempranos suyos, «El realismo» (1963) y «La realidad del pasado» (1969), me plantearé si el esquema realismo/antirrealismo de Dummett es realmente fructífero para el progreso de la metafísica[3]. También examinaré el conocido argumento que parece mostrar que el antirrealismo conduce a una contradicción.

## El realismo

Para muchos, el realismo es nuestra concepción tácita del mundo. Las montañas, los continentes y los planetas, por ejemplo, existen independientemente de nosotros. Aunque nosotros no hubiéramos existido, ellos existirían; aunque dejáramos de existir, seguirían existiendo; y su naturaleza no viene fijada por lo que nosotros podamos creer que ella es. Las montañas, los continentes y los planetas existen y lo hacen con total independencia de nosotros. El idealismo implica un rechazo radical de este realismo. Según el idealista, las montañas, los continentes y los planetas son entidades dependientes de la mente. Para tales entidades, ser consiste en ser percibido *(esse est percipi)*, ya sea por una mente humana o por la mente divina. Para el idealista, si no hubiera mentes, no habría montañas, continentes ni planetas. El antirrealismo de Dummett intenta ocupar una posición a medio camino entre el realismo y el idealismo. El mundo no es independiente de nosotros, ya que los enunciados acerca del mundo exterior no pueden ser verdaderos a menos que los podamos reconocer como tales. Pero la existencia de los planetas, por ejemplo, no dependería de que sean percibidos por alguna mente. Los planetas existirían aunque no hubieran existido las mentes.

## Demarcación de la disputa entre el realismo y el antirrealismo

Nuestra primera tarea es, por supuesto, aclarar lo que Dummett quiere decir con los términos «realista» y «antirrealista». Dummett declara que el realismo sobre los F, en el sentido deseado, no consiste en la doctrina de que hay F (donde los F son ciertas entidades cuyo estatuto está en cuestión), así como el antirrealismo tampoco consiste en el rechazo de dicha doctrina. Por lo tanto, la disputa realismo/antirrealismo sobre la existencia de los universales no sería un ejemplo de una disputa realismo/antirrealismo en el sentido de Dummett (véase el capítulo 3). La disputa realismo/antirrealismo de Dummett no se ocupa de la existencia de una clase de entidades cuestionada, sino de los rasgos de ciertas clases de enunciados (enunciados sobre otras mentes, sobre las matemáticas, el mundo físico, el pasado y el futuro, etc.). A Dummett le gusta citar la sentencia de Kreisel según la cual la mejor manera de entender el **platonismo** en matemáticas es viéndolo no como una doctrina sobre la existencia de objetos matemáticos, sino como una tesis sobre la objetividad de los enunciados matemáticos. El realismo, en el sentido de Dummett, sería también una manifestación de ese tipo de objetividad.

Donde la «clase problemática» [*disputed class*] se refiere a la clase de los enunciados relativos a un área respecto de la cual se da una disputa de realismo/antirrealismo, Dummett caracteriza al realismo como: «la creencia de que los enunciados de la clase en discusión poseen un valor de verdad objetivo, independientemente de los medios de que dispongamos para determinarlo: dichos enunciados serían verdaderos o falsos en virtud de una realidad que existe independientemente de nosotros»[4]. Por contraste:

> El antirrealista opone a ello la idea de que los enunciados de la clase en discusión deben ser entendidos sólo en referencia al tipo de cosas que consideramos evidencia para los enunciados de dicha clase [...] los significados de tales enunciados estarían directamente ligados a lo que nosotros consideremos evidencia para ellos, de tal manera que un enunciado de la clase en discusión, si es verdadero, sólo puede serlo en virtud de algo que nosotros deberíamos conocer y que debería valer como evidencia respecto de su verdad[5].

Por lo tanto:

> El realista y el antirrealista pueden perfectamente coincidir en que es objetiva la cuestión de si, respecto de un enunciado cualquiera de la clase en cuestión, los criterios que usamos para determinar si ese enunciado es verdadero han sido satisfechos o no: la diferencia entre ellos residiría en que, para el antirrealista, la verdad del enunciado sólo puede consistir en la satisfacción de dichos criterios, mientras que para el realista, el enunciado puede ser verdadero aunque no tengamos manera de reconocerlo como tal[6].

Más tarde escribe:

> la disputa puede surgir sólo para clases de enunciados respecto de las cuales ambas partes admiten que podría no existir evidencia a favor o en contra de un enunciado dado. Es esto, por lo tanto, lo que hace que la aceptación de la ley del tercio excluso para enunciados de una clase dada sea una prueba crucial para determinar si alguien adopta o no una posición realista respecto de enunciados de esa clase. El antirrealista no puede permitir que la ley del tercio excluso sea universalmente válida[7].

Consecuentemente, «La disputa [...] concierne a la noción de verdad que se adapte mejor a los enunciados de la clase en discusión; y esto significa que se trata de una disputa que atañe al tipo de *significado* que poseen estos enunciados»[8].

Cada clase de enunciados problemática suele ir acompañada de una clase de enunciados reductiva [*reductive class of statements*]. Así, la clase problemática formada por los enunciados sobre los objetos materiales suele ir acompañada de la clase (reductiva) de enunciados sobre las sensaciones. Los enunciados sobre las otras mentes suelen ir acompañados de enunciados sobre el comportamiento. Con la expresión «clase reductiva», Dummett quiere dar a entender que los enunciados de esa clase pueden ser entendidos independientemente de los enunciados de su respectiva clase problemática. En tales casos, la concepción antirrealista implica que la verdad de un enunciado de la clase problemática consiste en la verdad de algún enunciado, o conjunto de enunciados, de la clase reductiva[9].

Sin embargo, el antirrealismo no debería ser identificado con el reduccionismo: el reduccionismo no es condición suficiente ni necesaria del antirrealismo. No es necesaria, ya que no hay una clase reductiva que corresponda a la clase problemática de enunciados matemáticos (a saber, la de aquellos enunciados matemáticos que cuantifican un dominio infinito). La evidencia de un enunciado matemático consiste en una prueba o un cálculo. Sin embargo, la prueba es prueba *de* un enunciado matemático. Por lo tanto, la noción de prueba no es inteligible independientemente de la noción de enunciado matemático y, en consecuencia, no puede formar una clase reductiva para los enunciados matemáticos.

Consideraciones similares valen para los enunciados que se refieren al futuro o al pasado. Parte de la base evidencial para enunciados sobre el pasado son los recuerdos, pero todo recuerdo es un recuerdo de estados de cosas pretéritos, por lo que no puede ser caracterizado independientemente del pasado. Asimismo, parte de la base evidencial para enunciados sobre el futuro son las intenciones, pero toda intención es intención de hacer cosas en el futuro, por lo que no puede ser caracterizada independientemente del futuro.

Por otra parte, la razón de que el reduccionismo no sea condición suficiente del antirrealismo es que es posible suscribir una concepción realista respecto de la clase reductiva. En tal caso, la verdad de un enunciado de la cla-

se en discusión podría deberse a la verdad de un enunciado, o conjunto de enunciados, de la clase reductiva; pero como estamos siendo realistas respecto de la clase reductiva, también debemos serlo respecto de la clase en discusión.

Examinaremos a su debido tiempo la motivación que subyace al antirrealismo. Pero quizás resulte útil analizar con cierto detalle por lo menos dos casos de estudio del realismo/antirrealismo ofrecidos por Dummett —los referidos al carácter y al pasado— para tomarle el pulso al debate y a las distintas opciones que el realista y el antirrealista tienen a su disposición. Dummett describe el caso del carácter como uno en el que «muy pocas personas adoptarían sinceramente una actitud realista»[10]. Como veremos, se trata de una afirmación bastante discutible.

## Caso de estudio I: el carácter

El rasgo de carácter aquí elegido es el de la valentía. Al escoger su ejemplo, Dummett hace dos supuestos simplificadores. En primer lugar, supone que no hay vaguedad alguna en la aplicación del predicado «valiente». Cualquier reacción ante una situación peligrosa puede ser clasificada como valiente o como no valiente. En segundo lugar, supone que una acción valiente individual es suficiente para justificar la atribución de valentía al agente respectivo. Ninguno de estos dos supuestos es verdadero, pero, de todas maneras, Dummett considera que el punto no reviste mayor importancia, y se vale de estas simplificaciones con el solo propósito de ilustrar los aspectos esenciales de la disputa realismo/antirrealismo.

Ahora preguntémonos respecto de un señor Jones recientemente fallecido si era valiente o no. Supongamos que Jones llevó una vida protegida en una universidad y nunca tuvo que enfrentarse al peligro. No tenemos, pues, evidencia alguna a favor de «Jones era valiente» ni a favor de «Jones no era valiente». En este caso, declara Dummett, «Jones era valiente» será verdadero sólo si el condicional contrafáctico «si Jones se hubiera tenido que enfrentar a un peligro, habría actuado valientemente» es verdadero. De forma similar, «Jones no era valiente» será verdadero, si el condicional contrafáctico «si Jones se hubiera tenido que enfrentar al peligro, no habría actuado valientemente» es verdadero.

Como afirma Dummett, podríamos tener evidencia indirecta a favor de alguno de estos dos condicionales contrafácticos si la valentía estuviera asociada a otros rasgos de carácter, de cuya posesión por parte de Jones tuviéramos constancia. Pero supóngase que no disponemos de tal evidencia supletoria. En ese caso, no tendremos fundamento para suscribir ninguno de los dos condicionales contrafácticos mencionados.

Dummett hace entonces una aseveración adicional: un condicional contrafáctico no puede ser verdadero por sí solo [*barely true*]. Si es verdadero, lo

será en virtud de algún enunciado categórico (no condicional). Como dice Dummett, este principio es «intuitivamente convincente»[11]. Podemos ilustrar este principio en el caso de una propiedad disposicional como la solubilidad, para expresar la cual nada ser presta mejor que un contrafáctico. Un terrón de azúcar es soluble si, y sólo si, en caso de que fuera sumergido en agua, se derritiría. Y este contrafáctico es verdadero en virtud de algún hecho categórico sobre la estructura química del azúcar.

Todo lo anterior, suponemos, se mantiene en el territorio común al realista y al antirrealista. Pero ahora se bifurca el argumento. El antirrealista arguye que ninguno de los dos citados contrafácticos referidos a Jones es verdadero, ya que no hay enunciado categórico actual alguno sobre el comportamiento de Jones que justifique cualquiera de ellos. Pero si ninguno de los dos contrafácticos es verdadero, entonces «Jones era valiente» no será verdadero o falso. Por lo tanto, obtenemos el siguiente pensamiento típicamente antirrealista: «Jones era valiente» no es verdadero ni falso porque no tenemos evidencia que hable a favor o en contra de este enunciado, ni garantía alguna de llegar a adquirirla en el futuro.

Por su parte, el realista concede que la verdad de un enunciado contrafáctico debe estar basada en la verdad de un enunciado categórico, pero afirma que uno de los condicionales contrafácticos está basado en un enunciado categórico verdadero (pero desconocido) sobre el carácter de Jones. Así pues, podría suceder que el contrafáctico «si Jones se hubiera tenido que enfrentar a un peligro, habría actuado valientemente» fuera verdadero en virtud de la verdad categórica (desconocida por nosotros) «Jones era valiente».

El realista considera el contrafáctico verdadero en virtud de determinada verdad sobre el carácter de Jones, mientras que el antirrealista considera la valentía (o ausencia de valentía) de Jones algo fundado sobre determinada verdad acerca de su comportamiento (de manera que si no existe tal verdad sobre su comportamiento, no tenga sentido plantearse si Jones era valiente o no). El realista presupone que «Jones era valiente» es o verdadero o falso, aunque nosotros no podamos saber si es lo uno o lo otro. El antirrealista, en cambio, presupone que, puesto que no podemos determinar si es lo uno o lo otro, «Jones era valiente» no es ni verdadero ni falso.

De acuerdo con el realista, «Jones era valiente» debe «*ser* o verdadero o falso, ya que el carácter del sujeto en cuestión —entendido como un mecanismo interno que determina su comportamiento— tuvo que poseer la virtud de la valentía o carecer de ella»[12]. No obstante, añade Dummett, «sólo una persona de mucha ingenuidad filosófica adoptaría una concepción realista sobre los enunciados referidos al carácter»[13]. Bueno, ingenua o no, la concepción realista del asunto resulta bastante intuitiva. Solemos concebir las características psicológicas como algo latente que da pie al comportamiento. Nos comportamos como lo hacemos *a causa de* ciertos estados mentales (por ejemplo, elevamos la voz porque estamos furiosos). Nuestros estados mentales no son idénticos a nuestro comportamiento, sino más bien lo que los causa. Asimismo, tampoco vemos di-

ficultad alguna en la idea de que, debido al autocontrol o a las condiciones externas, otra persona podría tener determinado rasgo psicológico aunque éste nunca se manifieste en su conducta. En cambio, el antirrealismo distorsiona nuestra manera de concebir los estados mentales de los demás, y presupone un **conductismo** ingenuo y poco plausible. ¿Qué otra cosa, de no ser por una adhesión al conductismo, podría motivar la acotación antirrealista de enunciados categóricos relevantes para la fundamentación contrafáctica a aquellos que se refieren al comportamiento públicamente observable de Jones?

### El Jones desconocido

El profesor Jones llevó una vida protegida en su ciudad universitaria, flanqueado por los anaqueles de su estudio, sin aventurarse nunca a aquellos rincones de la ciudad donde merodeaban rufianes y carteristas. Jones nunca tuvo que enfrentarse al peligro. Nunca fue amenazado ni se hizo daño a sí mismo, ni tampoco tuvo que presenciar el que otros sufrieran daños. Nada en su comportamiento proporciona evidencia de la verdad de «Jones era valiente» o de la de «Jones no era valiente». Tampoco estamos en posesión de algún método que nos permita zanjar esta duda aplicando un procedimiento mecánico en un periodo de tiempo finito. Supongamos, además, que en el futuro tampoco aparecerá evidencia al respecto. Entonces, afirma Dummett, no es el caso que el enunciado «Jones era valiente» sea verdadero o falso. El realista, en cambio, dirá que «Jones era valiente» es una oración declarativa precisa e inequívoca en castellano: verdadera, si Jones tenía la cualidad de la valentía, y falsa en caso contrario. Puesto que Jones habría tenido o no esa cualidad, el enunciado «Jones era valiente» es verdadero o falso, incluso si no sabemos si es lo uno o lo otro. Nuestra evidencia respecto de un enunciado debería ser separada de la cuestión de la verdad o falsedad de éste.

## Caso de estudio II: el pasado

Según una manera de concebir los enunciados contingentes futuros, como, por ejemplo, «habrá una batalla naval mañana», estos enunciados no son considerados ni verdaderos ni falsos. Sin embargo, los filósofos que suscriben esta posición no lo hacen por consideraciones antirrealistas, sino porque creen que el futuro es irreal. No hay hechos futuros que hagan que nuestros enunciados sobre el futuro sean verdaderos o falsos ahora. No ha sido tradicional, en cambio, sostener una concepción análoga sobre los enunciados en pretérito, ya que, según la mayoría de las teorías sobre el tiempo, el pasado es real. (Véase el capítulo 5.) Dummett piensa, a pesar de ello, que un argumento antirrealista puede ser desarrollado para ambas clases de enunciados, aunque aquí nos limitaremos a discutir los enunciados sobre el pasado.

En su artículo «El realismo», Dummett esboza brevemente cómo podría presentarse una disputa realismo/antirrealismo sobre el pasado. Hay, presumiblemente, numerosos enunciados sobre el pasado a favor o en contra de los cuales no tenemos, ni posiblemente tengamos nunca, evidencia alguna (por ejemplo, «César desayunó huevos fritos en su trigésimo cumpleaños»). El realista insistirá en que tales enunciados, a menos que adolezcan de vaguedad, ambigüedad o ausencia de referente, son o verdaderos o falsos. El antirrealista, en cambio, subrayará que, dada la ausencia de evidencia presente o futura, tales enunciados no son ni verdaderos ni falsos.

En «La realidad del pasado», Dummett propone la siguiente estrategia para el realista. El realista podría justificar su comprensión de los enunciados acerca del pasado apelando a los nexos veritativos entre los enunciados conjugados en presente y los enunciados conjugados en pretérito. Si yo digo «Garrett está en su oficina», esto debe tener el mismo valor de verdad que el enunciado formulado dentro de un año «Garrett estaba en su oficina hace un año». Es imposible que estos enunciados tengan valores de verdad diferentes: de ahí el vínculo veritativo que los une. De acuerdo con Dummett, el realista debería suscribir la tesis de que «de la comprensión del vínculo veritativo, ejemplificado en un caso como éste, derivamos nuestra comprensión de lo que significa que un enunciado conjugado en pretérito [...] sea verdadero»[14].

Vale la pena destacar aquí tres aspectos. En primer lugar, Dummett presupone que, si fuéramos realistas sobre el pasado, podríamos hacerlo únicamente apelando a nexos veritativos que conectasen los enunciados conjugados en presente con los conjugados en pretérito. Pero ¿por qué un realista debería aceptar semejante cosa? Conceder eso equivaldría a conceder una idea típicamente antirrealista: que tratamos de comprender los enunciados conjugados en pretérito partiendo de la evidencia disponible en el presente o en futuro. Pero para un realista, «César desayunó huevos fritos en su trigésimo cumpleaños» es una oración enunciativa perfectamente clara y significativa, inteligible por sí misma para los hablantes del castellano. No requiere nexos veritativos que le confieran un valor de verdad determinado.

Un cierto patrón comienza a emerger: en el ejemplo de Jones, Dummett presuponía que los enunciados sobre otras mentes debían estar basados en enunciados sobre la conducta. Ahora presupone que los enunciados sobre el pasado deben estar basados en enunciados sobre la evidencia presente y futura. Ambas presuposiciones distorsionan nuestra comprensión de las otras mentes y del pasado. Ningún realista (ni ninguna otra persona, a decir verdad) debería aceptarlas.

En segundo lugar, el antirrealismo sobre el pasado es contraintuitivo. Damos por sentado que hay un pasado real y efectivo, y que o bien incluye el hecho de que César desayunó huevos fritos en su trigésimo cumpleaños o no lo incluye. ¿Cómo, entonces, podría dejar de ser verdadero o falso que César desayunó huevos fritos en su trigésimo cumpleaños?

En tercer lugar, hay una anomalía adicional. Hemos venido suponiendo que no hay evidencia presente o futura respecto del enunciado «César desayunó huevos fritos en su trigésimo cumpleaños». El antirrealista afirma que el enunciado expresado por medio de esta oración no es ni verdadero ni falso. Pero supongamos por un momento que sí tuviéramos evidencia de que César desayunó huevos fritos en su trigésimo cumpleaños. Entonces estaríamos justificados en concebir el correspondiente enunciado como verdadero o falso (ya que ahora sí tendríamos evidencia para afirmar su verdad). Pero ¿no resulta extraño que estos dos enunciados difieran tan marcadamente entre sí con respecto a la posesión de un valor de verdad determinado?

En el caso de los enunciados matemáticos, el antirrealismo es algo más plausible que en los casos anteriores. El antirrealista no tiene reparo en admitir que los enunciados matemáticos que se aplican a un dominio finito, por muy grande que éste sea, pueden ser concebidos como si poseyeran un valor de verdad definido. En cambio, no cree que se pueda atribuir a los enunciados que se extienden a un dominio infinito, y que, por consiguiente, no son susceptibles de pruebas ni de contraejemplos, un valor de verdad determinado. Así pues, el antirrealista admitiría que «el primer millardo de millardos de los números pares resulta de la suma de dos números primos» es o verdadero o falso, incluso si nadie lo ha verificado o falsado hasta ahora, ya que podría ser verificado o falsado en un tiempo finito usando un procedimiento mecánico de cálculo. Pero la conjetura de Goldbach —de que cada número par es la suma de dos primos— no podría ser, según él, considerada verdadera o falsa porque, en la actualidad, no tenemos para ella una prueba o un contraejemplo, ni la garantía de llegar a obtener una u otro alguna vez. Pero aquí por lo menos cabría hacer una distinción categorial entre aquellos enunciados a los cuales el antirrealista está dispuesto a otorgar un valor de verdad definido y aquellos a los que no está dispuesto a otorgárselo. El antirrealismo sobre el pasado no permite trazar una demarcación semejante.

## Motivación y balance del antirrealismo

Como ya habrá quedado de manifiesto, el proyecto antirrealista tiene su origen en consideraciones que tienen que ver con el significado de la verdad. Pero ¿de qué consideraciones estamos hablando? En su artículo temprano, Dummett decía más o menos lo siguiente: para el antirrealista «el significado de un enunciado está intrínsecamente conectado con aquello que consideramos evidencia a favor o en contra de ese enunciado»[15]. En su artículo posterior, se explaya un poco más. Dice que el antirrealista:

> sostiene que el procedimiento que nos permite captar el sentido de los enunciados de la clase problemática, así como el uso que subsecuentemente se hace de dichos enunciados, son tales, que de ello no podríamos derivar noción alguna de lo que significa que un enunciado

de esa clase sea verdadero independientemente del tipo de cosas que hemos aprendido a reconocer como lo que determina la verdad de dichos enunciados [...] Nada en la naturaleza misma del caso nos permitiría comprender lo que es para el enunciado ser verdadero independientemente de aquello que hemos aprendido a tratar como lo que establece su verdad: sencillamente no hay recursos con los que se nos hubiera podido mostrar eso[16].

Es difícil apreciar la fuerza de estas consideraciones. Tómese de nuevo la conjetura de Goldbach respecto de que cada número par es la suma de dos números primos. Ésta es una conjetura para la que, hasta el presente, no disponemos de prueba o contraejemplo alguno. Dado que los números pares forman una serie infinita, no hay, además, garantía de que alguna vez vayamos a obtener una u otro. Es, por consiguiente, una conjetura respecto de la cual se debaten el realista y el antirrealista. Algunos comentarios de Dummett dan la impresión de que el antirrealista debería afirmar que la oración «todo número par es la suma de dos primos» carece de significado. Pues si el significado está ligado a la evidencia, y no tenemos evidencia a favor o en contra de esta oración, ¿no se sigue que carece de significado? Sin embargo, ésta no es —por suerte, ya que es obvio que la oración sí posee un significado— la opinión ponderada de Dummett.

Lo que Dummett reclama no es que la oración carezca de significado, sino que no tenemos idea alguna de «lo que sería para tal enunciado ser verdadero independientemente del tipo de cosas que hemos aprendido a reconocer como lo que determina la verdad de enunciados de esa clase» (en este caso, independientemente de una prueba)[17]. Dada la ausencia de evidencia a favor o en contra, y dada la ausencia de garantía alguna de que vayamos a adquirirla alguna vez, no podemos pensar sobre la conjetura como algo que es verdadero o falso. En su artículo previo, Dummett había expresado esto diciendo que deberíamos considerar un enunciado problemático, tal como la conjetura de Goldbach, algo que no es ni verdadero ni falso. Sin embargo, esta aseveración fue en el artículo posterior sustituida por la de que no estamos justificados para afirmar que la conjetura de Goldbach debe ser verdadera o falsa (donde no estar justificado a afirmar no es equivalente a estar justificado a negar). Ésta es una modificación sutil pero importante. Sería obviamente absurdo sostener que la conjetura de Goldbach no es ni verdadera ni falsa, simplemente porque no tenemos en el momento actual una prueba o un contraejemplo de ella. Es como si tratáramos de probar que la conjetura de Goldbach es falsa sólo porque carecemos de una prueba o un contraejemplo. Sin embargo, ¡los resultados matemáticos no se obtienen tan fácilmente! En cambio, el negarse a afirmar que la conjetura de Goldbach sea verdadera o falsa no tiene la misma consecuencia absurda.

Y ¿qué decir de la motivación del antirrealismo en el campo de las matemáticas? ¿Por qué pensar que nuestra concepción de lo que significa para la conjetura de Goldbach el ser verdadera depende de nuestra posesión de una prueba de ella o de nuestra capacidad para reconocer una prueba cuando nos la topamos? Tales consideraciones sobre la evidencia parecen irrelevantes en

cuestiones de semántica. Una vez que logro captar el significado de «cada número par es la suma de dos números primos», entiendo que expresa una aseveración clara y definida sobre la realidad, como lo haría cualquier oración en castellano que fuera precisa e inequívoca. La oración «todo número par es la suma de dos primos» presenta de una manera definida la realidad (y, en particular, aquella parte de la realidad que consiste en la serie de los números pares). La oración es verdadera si la realidad es de esa manera; y falsa, si la realidad no es así. Esto capta con toda exactitud lo que está involucrado en mi comprensión de la conjetura de Goldbach, sin, para ello, tener que hacer referencia a la evidencia que yo pueda tener a favor o en contra de dicha conjetura. Estas observaciones parecen simples perogrulladas, pero sirven para desbaratar el proyecto antirrealista desde sus inicios. Pero si es así, estamos perfectamente justificados para considerar la conjetura de Goldbach verdadera o falsa, a pesar de que no tenemos idea de cuál sea su valor de verdad[18].

Dummett sin duda concibió el antirrealismo como un perfeccionamiento del positivismo lógico de Alfred J. Ayer[19]. El positivismo lógico fue una teoría del significado (literal) de las oraciones. De acuerdo con esta teoría, una oración enunciativa era literalmente significativa sólo si era o analítica (verdadera o falsa sólo en virtud de su significado) o empíricamente verificable. Alfred Ayer se vio a sí mismo como miembro de la larga tradición del empirismo británico que se retrotraía a Locke y como un verdugo de los dragones metafísicos. El positivismo lógico fue, ciertamente, una teoría **empirista**, que tuvo implicaciones para la significatividad [*meaningfulness*] de ciertas oraciones de nuestro idioma. Así, considérese la oración «todo está duplicando su tamaño». Esta oración no es ni analítica ni empíricamente verificable. Por principio, no podríamos tener evidencia sensorial a favor o en contra de esta oración (ya que no existe un punto focal neutral que nos permita verificarla). Si todo (incluyéndonos a nosotros mismos) estuviera duplicando su tamaño (exactamente al mismo tiempo), entonces no nos daríamos cuenta en absoluto del cambio respectivo. Por lo tanto, el positivismo lógico tendría que declarar que esta oración carece de significado. Y, sin embargo, es obvio que la oración tiene un significado. Sabemos exactamente lo que quiere decir.

Una supuesta ventaja del antirrealismo consistía en evitar la consecuencia de que una oración como «todo está duplicando su tamaño» careciera de sentido. La idea era que no nos sintiéramos justificados a afirmar categóricamente que la oración debía ser verdadera o falsa. Pero, a decir verdad, esto último es casi tan contraintuitivo como lo primero. La oración no carece de referente, ni tampoco es vaga o ambigua. Se trata de una oración enunciativa en castellano que hace una aseveración inteligible sobre la realidad. ¿Cómo podría la realidad no ser, ya como ella la describe, ya de otra manera? Con lo cual no nos quedaría más remedio que concebirla como verdadera o falsa. Quizás parezca admirable que Dummett se esfuerce en mantener vivo el espíritu de la gloriosa tradición empirista británica, pero no por ello deja de ser ésta una causa perdida[20].

### A. J. Ayer (1910-1989)

Sir Alfred Jules Ayer recibió su educación universitaria en Eton y Christ Church, Oxford. Obtuvo la cátedra Grote en el University College, de Londres, antes de ocupar la Wykeham de Lógica, de Oxford, en 1959. Personaje pintoresco y controvertido, tuvo un amplio círculo de amigos y dividió gran parte de su tiempo entre Londres y Oxford. Definitivamente, ha sido el único filósofo que ha logrado grabar un disco con Lauren Bacall. Uno de los más destacados filósofos británicos del pasado siglo, se hizo famoso aún siendo muy joven con la publicación en 1936 de su libro *Lenguaje, verdad y lógica*. En éste se afirmaba que las proposiciones de la metafísica carecen de sentido, ya que no son ni analíticas ni empíricamente verificables. Las proposiciones de la lógica y las matemáticas, e incluso todas las proposiciones a priori, eran catalogadas como analíticas. Y a las oraciones de la ética, tales como «la tortura es moralmente condenable», que no son ni analíticas ni empíricamente verificables, se las declaraba significativas sólo en virtud de que expresarían actitudes o emociones, en lugar de enunciar hechos éticos. Aunque Ayer posteriormente modificó las doctrinas radicales de su obra temprana, nunca renunció a su enfoque empirista.

## ¿Una refutación del antirrealismo?

La discusión precedente no ha sido especialmente amable con la polémica antirrealista. Desafortunadamente, las cosas van a empeorar. Además de lo contraintuitivo de las aseveraciones antirrealistas y su carencia de una motivación plausible, hay un poderoso argumento contra la coherencia del antirrealismo. El argumento se debió originalmente a F. B. Fitch, y fue publicado por primera vez en 1963[21].

Una consecuencia del antirrealismo es que no hay verdades incognoscibles. Si las hubiera, entonces un enunciado podría ser verdadero y, por ende, veritativamente determinado, aunque no hubiera posibilidad de que nosotros llegásemos jamás a reconocerlo como tal. Y es precisamente esta independencia de la mente con respecto a la realidad la que el antirrealista desea excluir. El realismo, en cambio, es plenamente consistente con la existencia de verdades incognoscibles.

Por lo tanto, podemos dar por sentado que el antirrealista está comprometido con el principio siguiente:

(KP) Para todos los enunciados p, si p es verdadero, entonces es posible conocer p.

Antes de formular el argumento de Fitch, necesitamos una premisa adicional. Aunque el antirrealista no puede tolerar las verdades incognoscibles,

debe admitir que algunas verdades no son conocidas. Esto se sigue simplemente del hecho de que no somos omniscientes. A diferencia de Dios, no conocemos todas las verdades que existen. Prescindiendo de algunas formulaciones ambiguas, el antirrealista Dummett no tiene empacho en admitir que hay muchas verdades para las que carecemos de evidencia (ya por el mero hecho de que nadie se haya molestado en recoger esta última, ya porque les llevaría a los seres humanos demasiado tiempo recopilarla). Por lo tanto, para el antirrealista, aunque no hay verdades incognoscibles, hay muchas verdades que desconocemos. Sea q una verdad desconocida dada.

La prueba de Fitch procede como sigue (donde «K» representa «es sabido que...», y «~» representa «no»):

| | |
|---|---|
| (1) q y ~Kq | Premisa |
| (2) Posiblemente K(q y ~Kq) | (1) ([KP]) |
| (3) Posiblemente (Kq y K~Kq) | (2) (distribución del conocimiento respecto de la conjunción, y principio de posibilidad) |
| (4) Posiblemente (Kq y ~Kq); | (3) (el conocimiento implica la verdad, y principio de posibilidad) |
| (5) (KP) es falso; | (4) *reductio* |
| (6) El antirrealismo es falso | (5) (el antirrealismo implica [KP]) |

Empecemos por (1). La premisa (1) es un supuesto que hace prácticamente todo el mundo. Expresa nuestra falta de omnisciencia. La premisa (2) se sigue de (1) por (KP) arriba (sustituyendo «p» por «q y ~Kq»). (2) nos dice que es posible saber (q y ~Kq). La premisa (3) se sigue de (2) por el principio de que el conocimiento es distributivo respecto de la conjunción, y el principio de posibilidad. Lo primero indica que quienquiera que conoce una conjunción conoce también cada miembro de ella: de «X conoce A y B» se puede inferir «X conoce A» y «X conoce B». Lo segundo indica que de «posiblemente R» y «R implica S» se puede inferir «posiblemente S».

La premisa (4) se sigue de la (3) por el principio de que el conocimiento implica la verdad: de «X sabe que A» podemos inferir la verdad de «A». Este principio se considera esencial al conocimiento, y es un aspecto capital por el que el conocimiento se distingue de estados «no factivos» [*non-factive states*] como la creencia (de «X cree que p» no podemos inferir «p»; X podría equivocarse)[22]. El paso de (3) a (4) requiere la aplicación del principio de posibilidad.

Pero (4) es imposible; por lo tanto, alguna premisa o principio previo debe estar errado. La premisa (1) es inobjetable; por otra parte, el principio de posibilidad, así como el principio de que el conocimiento es distributivo respecto de la conjunción, y el principio de que el conocimiento implica la verdad, parecen todos innegables; por consiguiente, la fuente de la contradicción debe ser (KP). Por lo tanto, (KP) es falso, y, como el antirrealismo implica (KP), el antirrealismo también es falso.

Aunque no tengo reparos en tratar la prueba de Fitch como una refutación del antirrealismo exitosa, debo advertir que algunos filósofos critican como paradójico el razonamiento de Fitch y se esfuerzan por descubrir alguna falacia en él[23]. Aceptar como válida la prueba de Fitch equivale a aceptar que si algunas verdades son desconocidas, entonces algunas verdades son incognoscibles. Por **conversión** [*contraposition*] esto equivale al principio de que si todas las verdades son cognoscibles, entonces todas las verdades son conocidas. Quizás se considere este principio contraintuitivo, más allá de la disputa realismo/antirrealismo[24]. Sin embargo, no hay un consenso sobre lo que estaría errado en la prueba de Fitch, y hasta que tal consenso emerja, es razonable aceptar la prueba de Fitch como una refutación de (KP) y, por lo tanto, como una refutación del antirrealismo.

## Observaciones finales

El proyecto antirrealista de Dummett no salió bien parado de nuestra discusión: se reveló como contraintuitivo, inadecuadamente moderado y refutable a través de la prueba de Fitch. Esta conclusión pesa negativamente en el proyecto más amplio de Dummett de colocar la teoría semántica en el centro de la filosofía. Dummett trató de delinear un nuevo sentido en el que el mundo (o algún aspecto de él) podría no ser independiente de nosotros y de nuestra capacidad para recolectar evidencia. Aunque no haya tenido éxito, quizás otros corran con más suerte[25].

## Preguntas para el estudio

- ¿Qué implica ser realista en un área concreta?
- ¿Cómo caracterizarías el antirrealismo de Dummett?
- ¿Es plausible afirmar que «Jones era valiente» no es ni verdadero ni falso?
- ¿Se te ocurre alguna motivación válida para ser antirrealista sobre un tema cualquiera?
- ¿Cómo podría un antirrealista responder a la prueba de Fitch?

### Notas

1. La condición (ii) nos permite ser realistas sobre los trenes, los aviones y los automóviles. Aunque éstos no existirían si nosotros no hubiéramos existido, seguirían existiendo aunque nosotros dejáramos de existir. La condición (iii) nos permite ser realistas acerca de los estados mentales. Aunque es cierto que, si yo no hubiera existido, tampoco lo habría hecho mis estados mentales, y que, si yo dejara de existir, también dejarían de existir mis estados mentales, la naturaleza de mis estados mentales no está fijada por la naturaleza que yo les atribuya. Como podríamos estar equivocados sobre la naturaleza de nuestros estados mentales, el realismo sobre los estados mentales es una

opción abierta.

2. Según una teoría correctiva [*error theory*] referida a los F, no hay F, y, por lo tanto, los enunciados sobre los Fs son falsos (erróneos). Según el expresivismo sobre los F, las oraciones declarativas que contienen el término «F» no se usan para formular enunciados, sino que tienen alguna otra función (como, por ejemplo, la de expresar actitudes o emociones). El expresivista generalmente está de acuerdo con el teórico del error en que no hay F.
3. M. A. E. Dummett, «Realism» y «The Reality of the Past», en la colección de sus artículos *Truth and Other Enigmas* (Cambridge, Mass.: Harvard University Press) 1978, pp. 145-165 y 358-374.
4. «Realism», *op. cit.*, p. 146.
5. Ibíd., p. 146.
6. Ibíd., p. 147.
7. Ibíd., p. 155. La ley del tercio excluso dice que, para todo P, P o no-P. Esta ley es distinta del principio de bivalencia (el principio de que todos los enunciados en un área determinada son verdaderos o falsos), aunque está relacionada con él. Sólo el segundo es el blanco primordial del antirrealista.
8. Ibíd., p. 146.
9. Aunque el propósito de Dummett es resolver cuestiones de metafísica, el antirrealista suele clasificar una clase de enunciados como «problemática» basándose en razones epistémicas. Es debido a que nuestro acceso, por ejemplo, a los objetos materiales, a otras mentes o al pasado es supuestamente «indirecto», por lo que los enunciados referidos a tales dominios (a diferencia, por ejemplo, de los enunciados sobre datos sensoriales, la conducta y los rasgos presentes) se consideran problemáticos. Adicionalmente, Dummett presupone que los enunciados sobre la experiencia sensorial pueden ser comprendidos sin hacer referencia a objetos materiales, y que los enunciados sobre el comportamiento humano pueden ser comprendidos sin hacer referencia a los estados mentales. Ambas presuposiciones son discutibles, y hasta se puede argüir que falsas. Véanse J. Cook, «Human Beings», en P. Winch (ed.), *Studies in the Philosophy of Wittgenstein* (Nueva York: Routledge) 1969, pp. 117-151, y J. McDowell, «Criteria, Defeasibility and Knowledge», en J. Dancy (ed.), *Perceptual Knowledge* (Oxford: Oxford University Press) 1988, pp. 209-219.
10. «Realism», *op. cit.*, p. 148.
11. Ibíd., p. 148.
12. Ibíd., p. 150.
13. Ibíd., p. 150.
14. «The Reality of the Past», *op. cit.*, p. 363.
15. «Realism», *op. cit.*, p. 162.
16. «The Reality of the Past», *op. cit.*, p. 362.
17. Ibíd., p. 362.
18. Al criticar la posición del antirrealismo, no estoy presuponiendo el principio de bivalencia. Muchas oraciones (por ejemplo, las que son vagas o ambiguas) no son ni verdaderas ni falsas. Lo que está en discusión son las razones del antirrealista para rechazar la bivalencia.
19. Véase A. J. Ayer, *Language, Truth and Logic* (Harmondsworth: Penguin) 1976.
20. Hay, por supuesto, mucho más que decir al respecto. No puedo hacer aquí otra cosa que remitir al lector interesado a tres artículos de Alex Miller, que revelan en detalle los errores y falacias de la posición antirrealista. Véanse su «The Significance of Se-

mantic Realism», *Synthese,* vol. 136, 2003, pp. 191-217; «What is the Manifestation Argument?», *Pacific Philosophical Quarterly,* vol. 83, 2002, pp. 352-383, y «What is the Acquisition Argument?» en A. Barber (ed.), *Epistemology of Language* (Oxford: Oxford University Press) 2003, pp. 459-494.

21. F. Fitch, «A Logical Analysis of Some Value Concepts», *The Journal of Symbolic Logic,* vol. 28, 1963, pp. 135-142.
22. Un estado $\Phi$ es factivo si, y sólo si, «$\Phi$ p» implica «p».
23. Véase, por ejemplo, D. Edington, «The Paradox of Knowability», *Mind,* vol. 94, 1985, pp. 557-568.
24. Por ejemplo, ¿no habría podido Dios crearnos con mentes increíblemente poderosas, capaces de aprehender cualquier verdad del universo, sin que ello fuera inconsistente con el hecho de que eligiésemos no investigar cada asunto? Esta aparente posibilidad queda descartada por la prueba de Fitch.
25. Por ejemplo, Crispin Wright ha explorado recientemente nuevas formas de oponerse a la presuposición del realista de que hay asuntos que se constituyen independientemente de nosotros. Véase, por ejemplo, su obra *Truth and Objectivity* (Cambridge, Mass.: Harvard University Press) 1992.

## Bibliografía adicional comentada

A. J. Ayer, *Language, Truth and Logic* (Harmondsworth: Penguin) 1976, especialmente la Introducción de Ayer y el capítulo 1. El libro de Ayer es una presentación clásica del positivismo lógico, doctrina que se puede considerar precursora (cercana) del antirrealismo.

M. A. E. Dummett, «Realism» y «The Reality of the Past», en la colección de sus artículos *Truth and Other Enigmas* (Cambridge, Mass.: Harvard University Press) 1978. Ambos trabajos son bastante difíciles, pero el primero es el más accesible de los dos.

D. Edington, «The Paradox of Knowability», *Mind,* vol. 94, 1985. Una discusión de nivel avanzado sobre la paradoja de Fitch.

J. McDowell, «Criteria, Defeasibility and Knowledge», en J. Dancy (ed.), *Perceptual Knowledge* (Oxford: Oxford University Press) 1988. Una fina crítica de los fundamentos epistemológicos del antirrealismo. Escrito con claridad, pero exigente.

C. Wright, *Realism, Meaning and Truth,* 2.ª edición (Oxford: Basil Blackwell) 1993. Todos los artículos de esta colección tienen que ver con el debate realismo/antirrealismo, pero son demasiado difíciles para el principiante. Sin embargo, la Introducción escrita por Wright es, de hecho, una genuina introducción al debate y digna de ser recomendada al aprendiz entusiasta.

## Recursos en Internet

B. Brogaard y J. Salerno (2004) «Fitch's Paradox of Knowability», *The Stanford Encyclopedia of Philosophy (edición de verano de 2004),* Edward N. Zalta (ed.). Bajado el 31 de mayo de 2006 de <http://plato.stanford.edu/archives/sum2004/entries/fitch-paradox>.

E. Craig (1998) «Realism and Antirealism», en E. Craig (ed.), *Routledge Encyclopedia of Philosophy.* Bajado el 31 de mayo de 2006 de <http://www.rep.routledge.com/article/N049>.

A. Miller (2005) «Realism», *The Stanford Encyclopedia of Philosophy (edición de otoño de 2005),* Edward N. Zalta (ed.). Bajado el 31 de mayo de 2006 de <http://plato.stanford.edu/archives/fall2005/entries/realism>.

B. Taylor (1998, 2004) «Dummett, Michael Anthony Eardley», en E. Craig (ed.), *Routledge Encyclopedia of Philosophy.* Bajado el 31 de mayo de 2006 de <http://www.rep.routledge.com/article/DD083>.

# Glosario

**a priori:** Se refiere a la forma en que una oración o enunciado es conocido. Una oración es conocida a priori si comprender la oración es suficiente para determinar su valor de verdad. «Todos los solteros no están casados» es una oración conocida a priori; pero «el agua hierve a cien grados Celsius» no lo es. La noción de a priori debería distinguirse de las de necesidad y analiticidad (la verdad en virtud del significado), aunque está relacionada con ellas. El conocimiento que no es a priori es a posteriori o empírico.

**abstracto:** Esta palabra tiene muchos significados, pero hay dos que aquí nos interesan especialmente. Cuando los filósofos describen los universales o los números como objetos abstractos, se refieren con «abstracto» al hecho de que estos objetos no están en el espacio y en el tiempo. Pero cuando los teóricos de los tropos describen los tropos como particulares abstractos, utilizan la palabra «abstracto» como sinónima de «sutiles» o «vagos». El tropo de la rojez de una bola de billar no es concebido como si estuviera fuera del espacio y el tiempo.

**accidental/esencial:** Esta distinción proviene de Aristóteles. Las propiedades de un objeto pueden dividirse en aquellas que son esenciales y aquellas que son accidentales. Si F es una propiedad accidental de *x*, entonces, aunque *x* sea F, podría no haber sido F. En cambio, si G es una propiedad esencial de *x*, entonces *x* no habría podido no ser G. En la jerga sobre los mundos posibles, si F es una propiedad accidental de *x*, entonces hay mundos posibles en los que *x* existe pero no es F; y si G es una propiedad esencial de *x*, entonces no hay mundo alguno en el que *x* exista pero no sea G. W. V. Quine se hizo célebre por tratar de desacreditar esta distinción, a la que denominó «insidiosa». (Véase el artículo de Quine «Reference and Modality», en su *From a*

*Logical Point of View* [Nueva York: Harper & Row], 1963.)

**analítico:** Se dice que una oración es analítica cuando es verdadera (o falsa) únicamente en virtud de su significado. Ejemplos de oraciones analíticas son «todos los solteros no están casados», «todas las solteras son mujeres», «todos los triángulos tienen tres lados». Una oración no analítica, o sintética, es verdadera (o falsa) en virtud de su significado y de los hechos del mundo. Ejemplos de oraciones no analíticas son «todos los solteros usan pantalones», «todas las solteras son infelices», «el triángulo es mi objeto geométrico favorito». En su famoso artículo de 1952, «Two Dogmas of Empiricism», el filósofo de Harvard W. V. Quine cuestionó el linaje y la importancia de la distinción analítico/sintético. (El artículo está incluido en su compilación *From a Logical Point of View* [Nueva York: Harper & Row], 1963.)

**antecedente:** En un condicional de la forma «si P entonces Q», P es el antecedente (y Q el consecuente).

**aseveraciones causales generales:** Son aquellas aseveraciones causales que se refieren a tipos de eventos u ocurrencias. Así, «el VIH es el causante del sida» es un ejemplo de una aseveración causal general. Puedo pronunciar esta oración sin tener una persona determinada en mente.

**aseveraciones causales singulares:** Son aseveraciones causales que se refieren a eventos particulares e irrepetibles. Así, «el ataque histérico de María causó la muerte de Guillermo» es un ejemplo de una aseveración causal singular.

**cambios McTaggart:** Son cambios que afectarían a los objetos meramente en virtud del paso del tiempo (por ejemplo, el cambio consistente en ser un año más viejo). Shoemaker propone que se descarten los cambios de este tipo, ya que, de otra manera, sería imposible, por definición, que existiera el tiempo sin el cambio. En nuestro mundo, los cambios McTaggart no tienen incidencia causal alguna, por lo que, en una acepción de «cambio», no constituyen cambios genuinos.

**circuitos causales:** Los circuitos causales [*causal loops*] son una posibilidad exótica. Si viajar al pasado fuera posible, un hombre podría viajar al pasado, preñar a su madre y convertirse en padre de sí mismo. La existencia de ese hombre representaría un circuito causal. Cada evento de un circuito es causado por otro evento del circuito, pero la existencia del circuito entero carece de causa.

**condicional contrafáctico:** Un condicional contrafáctico es un condicional con un antecedente falso que enuncia lo que sería el caso si el antecedente hubiera sido verdadero. Así, aunque yo no lance un ladrillo contra la ventana, puedo decir la verdad siguiente: «si yo hubiera arrojado un ladrillo contra la ventana, la ventana se habría hecho añicos». Todos usamos y entendemos los condicionales contrafácticos, pero hay gran discusión acerca de la lógica que subyace a los mismos.

**condicional de indicativo:** El condicional de indicativo [*indicative conditional*] es diferente del condicional contrafáctico. El condicional de indicativo meramente enuncia lo que, dada la verdad de alguna condición antecedente, sucederá o ha sucedido. Así, «si Guillermo asiste a la fiesta, se formará un escándalo» y «si hay huellas en la

arena, Federico estuvo aquí anoche» son dos condicionales de indicativo.

**condicional material:** La frase «condicional material» es un término técnico de la lógica, representado por «⊃». Una oración «P ⊃ Q» es falsa, si y sólo si P es verdadera y Q falsa; de otro modo, es verdadera. En otras palabras, «P ⊃ Q» es equivalente a «No-p o Q». Está fuera de duda que los condicionales contrafácticos no son materiales. (Si lo fueran, todos los contrafácticos serían verdaderos, cosa que no es cierta.) Algunas personas piensan que los condicionales indicativos son condicionales materiales. Pero esto es problemático. ¿Es verdadero un condicional indicativo del castellano ordinario cada vez que su antecedente es falso? ¿Es verdadero el enunciado «si 2 + 2 = 5, entonces yo soy chino»?

**condicional:** Un condicional es cualquier oración de la forma «si P, entonces Q», donde P es el antecedente y Q el consecuente.

**condiciones necesarias y suficientes:** La finalidad del análisis conceptual ha sido tradicionalmente la producción de condiciones necesarias y suficientes para la aplicación de un concepto dado. El ejemplo más célebre es el análisis platónico del conocimiento en términos de creencia verdadera justificada (*Teeteto* 201c-210d). Este análisis presupone que los conceptos de creencia, verdad y justificación pueden ser comprendidos independientemente del concepto de conocimiento. El análisis intenta, por lo tanto, proporcionar un análisis reductivo del concepto de conocimiento.

**conductismo:** Hay muchas variedades de conductismo [*behaviourism*], pero su idea central es que la mente puede ser reducida a, o identificada con, la conducta o las disposiciones de conducta (donde la conducta es entendida en términos no mentalistas). Sin embargo, dicha reducción o identificación viola un dato del sentido común, a saber, que los estados y eventos mentales causan la conducta y las disposiciones de conducta. Si mi picazón es lo que hace que yo me rasque, no puede ser identificada con el rascarse mismo (puesto que nada es causa de sí mismo). Además, el conductismo describe equivocadamente la forma en que conocemos nuestros propios estados mentales. Sé que tengo picazón porque lo siento, no porque observo la manera en que me comporto.

**contingente:** Una oración es contingente si es verdadera en ciertas condiciones posibles (o mundos posibles), y falsa en otras. Así, «llovió en Edimburgo el 1º de enero de 2006» es contingente: es verdadera, pero habría podido ser falsa. Una oración es no contingente, o necesaria, si, o bien es verdadera en cualquier circunstancia posible, o bien es falsa en cualquier circunstancia posible. Algunos suelen hablar también de seres contingentes (como tú o yo) y de seres necesarios (como Dios o el número 2).

**continuidad mental y/o corporal:** Algunos filósofos piensan que la identidad de una persona a través del tiempo puede ser entendida en términos de continuidad corporal, mientras que otros prefieren entender la identidad personal en términos de continuidad psicológica. Según la concepción somática, continúo existiendo tanto tiempo como exista mi cuerpo corporal viviente. Según la concepción psicológica, continúo existiendo tanto tiempo como exista el flujo de mi vida mental (mis creencias, recuerdos, esperanzas, temores, etc.). Los defensores de la concepción psicológica sostienen que

su concepción es la que mejor se adapta a nuestro concepto de persona (esto es, al concepto de un cierto tipo de entidad psicológica).

**conversión** [*contraposition*]: Cualquier condicional es lógicamente equivalente a su converso [*contrapositive*). Es decir, «si p, entonces q» es equivalente a «si no-q, entonces no-p». (Nótese que «si p, entonces q» no es equivalente a «si no-p, entonces no-q».)

**cuantificadores:** Se trata de palabras que nos indican qué proporción o cantidad de cosas tienen determinada propiedad. Así, como respuestas a la pregunta «¿cuántos F son G?» pueden entenderse todas las oraciones que se enumeran a continuación: todos los F son G; algunos F son G; casi todos los F son G; muchos F son G; pocos F son G; al menos 24 F son G; ningún F es G, etc. El desarrollo de la lógica cuantificacional por Frege en el siglo XIX representó un gran avance respecto de los sistemas de lógica previos.

**deíctico:** Una palabra es deíctica [*indexical*] si, y sólo si, su referencia está determinada (en parte) por el contexto en el que se pronuncia. Así, una expresión como «yo» es deíctica porque su referencia está determinada por la identidad de quien la pronuncia; «aquí» lo es, porque su referencia está determinada por el lugar del emisor; «ahora» lo es, porque su referencia está determinada por el momento en que se pronuncia. «Yo», «aquí» y «ahora» son términos referenciales, pero tienen el rasgo curioso de ser inmunes a una carencia de referente o a un error denotativo.

**empirista:** Los filósofos empiristas clásicos fueron Locke, Berkeley y Hume. La tesis central del empirismo es que el conocimiento sustancial del mundo sólo puede ser obtenido por vía de los cinco sentidos. La razón puede, sin duda, proporcionar conocimiento (a saber, de las verdades a priori de la lógica y de la matemática), pero tal conocimiento no sería sustancial. Los racionalistas sostienen, por el contrario, que la razón puede proporcionar conocimiento sobre el mundo. Un ejemplo obvio de un argumento racionalista es el argumento ontológico, que intenta probar la existencia de Dios apelando únicamente a la razón.

**estructura lógica:** El interés contemporáneo en la estructura lógica de las oraciones del lenguaje natural fue estimulado por los trabajos de Frege, Wittgenstein y Russell a finales del siglo XIX y comienzos del XX. Algunos pensaron que comprender la estructura lógica del lenguaje revelaría la estructura del mundo. Pero, incluso prescindiendo de esta aspiración, hay razones para interesarse en cuestiones de estructura lógica. Hacernos una idea clara de la lógica de nuestro lenguaje puede ayudarnos a evitar una ontología exagerada (véase la crítica de Russell a Meinong) o permitirnos explicar inferencias que de otra manera nos dejarían perplejos (por ejemplo, por qué «Guillermo corrió rápido» implica «Guillermo corrió»).

**expresiones de ocurrencia reflexiva:** La expresión «de ocurrencia reflexiva» [*token reflexive*] fue acuñada por el filósofo de la física alemán Hans Reichenbach. Las expresiones de ocurrencia reflexiva [*token-reflexive expressions*], también denominadas «términos deícticos» [*indexicals*], son palabras (como «yo», «tú», «aquí», «allí», «ahora», «hoy», «ayer», etc.) cuya referencia, cuando son usadas, viene determinada por factores contextua-

les. Dichos factores contextuales son, normalmente, personas (en «yo», «tú»), lugares (en «aquí», «allí») o momentos (en «ahora», «hoy», «ayer»). Cuando se trata de emisiones en primera persona, el factor contextual relevante es el emisor; cuando se trata de emisiones que contienen un «aquí», es el lugar de la emisión, y así sucesivamente. Esto explica por qué «yo», puesto en boca mía, se refiere a mí, pero puesto en boca tuya, se refiere a ti. Si cambia el contexto, también cambia el referente.

**fundamental:** A todos nos interesa la expresión «fundamental» [*fundamental*] de la manera en que aparece en oraciones de la forma «X es más fundamental para Z que Y». La forma comparativa «más fundamental» es inusual. En su discusión, McTaggart no usa «fundamental» en el sentido de «esencial». Parece significar, con ese término, «esencial y completo». Así, aunque X e Y sean ambos esenciales a Z, X será más fundamental, si los hechos-X agotan los hechos-Z, pero los hechos-Y no lo hacen. De este modo, McTaggart piensa que la serie A es más fundamental para el tiempo que la serie B.

**identidad cualitativa:** Este sentido de «identidad» está expresado en oraciones como «ellos son gemelos idénticos» y «ambos tenemos el mismo coche», etc. Esta última oración se refiere no a un solo coche, sino a dos, fabricados por la misma empresa. La identidad cualitativa no se ajusta a la ley de Leibniz: si A y B son gemelos idénticos, no es el caso que todo lo que es verdadero de A es también verdadero de B, y viceversa. Por ejemplo, los gemelos difieren en el momento exacto de su nacimiento y en sus trayectorias temporales posteriores.

**identidad numérica:** Este sentido de «identidad» es el que expresan oraciones como «agua es $H_2O$», «Superman es Clark Kent», «2 es la raíz cuadrada positiva de 4». Cada una de estas oraciones involucra exactamente una entidad, la cual es descrita o nombrada de diversas formas. Para la identidad numérica se cumple la ley de Leibniz. Esto es, si «A es B» expresa una identidad numérica, entonces todo lo que es verdadero de A también es verdadero de B, y viceversa.

**indeterminista:** Según algunas teorías, los fenómenos cuánticos son esencialmente indeterminados (indeterministas). Esto es, lo que sucede en un momento dado en el nivel cuántico podría no estar completamente determinado por lo que sucedió en periodos de tiempo anteriores. Cierto resultado podría ser sólo probable, sin estar determinado. Se suele considerar a dicho indeterminismo una consecuencia del principio de incertidumbre de Heisenberg.

**ley de Leibniz:** Es la ley de la lógica que establece que si A es idéntico a B, entonces cualquier propiedad de A es una propiedad de B, y viceversa. Algunas veces se la formula sin hacer referencia a propiedades, así: si A es B, entonces cualquier cosa verdadera de A es verdadera de B, y viceversa. Esta ley debe ser distinguida del principio de la identidad de los indiscernibles: si A y B comparten todas sus propiedades, entonces A es B.

**leyes de la naturaleza:** Un ejemplo estándar de una ley de la naturaleza es: todos los metales se dilatan al calentarse. Pero ¿qué son las leyes de la naturaleza? Algunos filósofos (los humeanos) piensan que las leyes de la naturaleza son simplemente regularidades bien

establecidas. Otros (los antihumeanos) piensan que las leyes implican algún tipo de necesidad que explica las regularidades observadas. Una ley genuina nos indica lo que *debe* ocurrir.

**lógicamente necesario:** Para cualesquiera proposiciones P y Q, se dice que P es lógicamente necesaria para Q si y sólo si es lógicamente imposible que Q sea verdadera y P falsa (esto es, si, y sólo si, Q implica P). Si P es necesaria para Q, Q es suficiente para P. Es lógicamente imposible que Q sea verdadera y P falsa si, y sólo si, la conjunción <Q y no-P> es contradictoria. Los teístas que propugnan la solución del «mal necesario» al problema del mal sostienen que una conjunción como <benevolencia sin sufrimiento> es contradictoria.

**modal:** Se refiere a la posibilidad y a la necesidad. Las oraciones modales son aquellas de la forma: posiblemente P; necesariamente P; A habría podido ser F; A es necesariamente G; B no puede ser G, etc. Las aseveraciones modales tienen diferente fuerza, dependiendo de la modalidad respectiva. Así, «yo no puedo levantar este coche» se refiere a una imposibilidad física (esto es, el que yo levante ese coche es incompatible con hechos sobre mi constitución física y con las leyes de la naturaleza). En cambio, «yo no puedo a la vez levantar y no levantar ese coche» se refiere a una imposibilidad lógica (el que yo simultáneamente levante y no levante el coche es incompatible con las leyes de la lógica). Hay, asimismo, otros tipos de modalidades (por ejemplo, las legales: «usted no puede estacionar aquí»). Nótese también que podemos distinguir entre oraciones modales *de dicto* y oraciones modales *de re.* En la oración *de dicto* «necesariamente: 2 + 2 = 4», la necesidad es predicada de una oración o proposición. En la oración *de re* «Sócrates es necesariamente humano», la necesidad o la humanidad esencial son predicadas de un objeto no proposicional: Sócrates.

**nexos veritativos:** Los nexos veritativos [*truth-value links*] se producen cada vez que dos lenguajes o modos de hablar se relacionan sistemáticamente entre sí. Por ejemplo, las oraciones de primera y tercera persona están sistemáticamente relacionadas entre sí. «Yo soy calvo» dicho por la persona X es verdadero si y sólo si «X es calvo», pronunciado por cualquier otra persona, es verdadero. O, de nuevo, «hoy está lloviendo», proferido en el día 1, es verdadero si y sólo si «ayer llovió», proferido en el día 2, es verdadero. Nótese que la equivalencia en el valor de verdad es necesaria. No es casual que las oraciones de cada par tengan el mismo valor de verdad.

**no hay acción a distancia temporal:** Es el correlato temporal de «no hay acción a distancia espacial». Abandonar este principio de temporalidad equivale a sostener que A, que está en $t_1$, puede producir B, que está en $t_2$, sin que haya algún evento después de $t_1$ y antes de $t_2$ suficiente para producir B. Así como muchos creen que es posible la acción a distancia espacial, Shoemaker piensa que es también posible la acción a distancia temporal.

**no ramificada:** Una relación se ramifica [*a relation branches*], cuando se da entre una cosa, por un lado, y dos o más cosas que son posteriores en el tiempo, por el otro. La fisión de una ameba es un ejemplo, tomado de la vida real, de una continuidad física que se despliega ramificadamente. La continuidad psicológica también puede adoptar una forma ramificada

[*branching form*], como muestra el caso de la «fisión» estudiado en el capítulo 8. Debido a que una cosa no puede ser idéntica a otras dos, algunas formas de entender los criterios de la identidad estipulan que sólo se da identidad allí donde está ausente un tipo relevante de ramificaciones.

**omnipotente:** Dios es considerado todopoderoso u omnipotente. Esto es, Dios es capaz de producir cualquier estado de cosas lógicamente posible. Así, no es un argumento en contra de la omnipotencia de Dios el que no pueda hacer que mi mesa sea a la vez redonda y cuadrada, ya que esto no es un estado de cosas posible. Algunos han creído ver en esto un motivo de paradoja: ¿Puede Dios crear una piedra tan pesada que no pueda levantar? Si se responde que sí, Dios parece carecer de omnipotencia; si se responde que no, de nuevo parece no tenerla. Pero se trata de una pseudoparadoja. No hay contradicción alguna en la idea de que uno de nosotros cree una piedra tan pesada que no pueda levantarla, pero la descripción de esta tarea como la tarea de un ser omnipotente es contradictoria y lógicamente imposible. Por lo tanto, no refuta la omnipotencia divina el que Dios sea incapaz de crear una piedra tan pesada que no pueda levantarla.

**omnisciencia:** Es tradicional concebir a Dios como un ser que todo lo sabe. Esto es, para cualquier proposición verdadera p, Dios sabe que p. Algunos han cuestionado la corrección de este aserto. Si (como algunos creen) Dios está fuera del tiempo, ¿puede saber que ahora son las cuatro de la tarde? ¿Acaso no se requiere estar dentro del tiempo para captar una verdad como ésa? O, de nuevo: sé que estoy cansado. ¿Puede Dios captar esta verdad, o sólo yo puedo hacerlo? Quizás haya que esperar por nuevos desarrollos en la filosofía del lenguaje para responder a estas preguntas.

**platonismo:** Platón sostuvo que las formas (las verdaderas realidades) existen fuera del espacio y del tiempo. Ser platónico acerca de cierto rango de entidades es, por lo tanto, sostener que esas entidades existen fuera del espacio y el tiempo. (Como se observó en el capítulo 9, Dummet adopta una postura algo diferente sobre la idea que subyace al platonismo.)

**proyección:** El proyectivismo [*proyectivism*] respecto de una propiedad determinada F es la concepción de que concebimos erróneamente F como una propiedad es de las cosas del mundo exterior, cuando en realidad dicha propiedad es generada en nuestras propias mentes y proyectada hacia el mundo. Así, podemos describir una situación como temible, pero eso es sólo porque reaccionamos a ella de una manera determinada. Otro ejemplo: algunos filósofos piensan que el mundo como tal puede ser vago; pero esto último parece ser una proyección de nuestros conceptos (vagos) a una realidad (precisa). Hume es la fuente clásica del proyectivismo. Habló de «la propensión de la mente a diseminarse sobre los objetos externos» (*Tratado*, 1.3.14, párr. 24).

**reducción:** En el debate sobre el tiempo, ¿qué significa decir que las verdades temporalmente contextualizadas [*tensed truths*], formuladas en el vocabulario de la serie A, pueden ser reducidas a verdades intemporales [*tenseless truths*], formuladas en el vocabulario de la serie B? Aunque el teórico B puede admitir verdades y hechos temporalmente contextualizados, además de verdades y hechos

intemporales, no puede verlos como si tuvieran el mismo estatus. Para el teórico B, la realidad es intemporal: la descripción correcta de ésta no requiere verdades temporalmente contextualizadas. Por ejemplo, aunque sea cierto que la muerte de Hitler sucedió en el pasado, y sea un hecho que esto es así, ello se debe únicamente a que cualquier emisión presente de la frase «la muerte de Hitler sucedió en el pasado» se ha dado con posterioridad a la muerte de Hitler. Según la teoría B, las verdades intemporales son últimas e irreductibles, mientras que las verdades temporalmente contextualizadas son dependientes y eliminables.

**reductivo:** Palabras como «reductivo» y «reduccionismo» son términos técnicos de la filosofía usados con distintos significados. Parfit sostiene la tesis de que somos reduccionistas sobre los F si pensamos que la realidad puede ser completamente descrita sin hacer referencia alguna a los F. Muchos son reduccionistas, en este sentido, sobre las entidades sociales, tales como las naciones o los comités. Así, aunque haya verdades sobre comités (por ejemplo, «el comité aprobó unánimemente la designación de Smith»), una descripción de la realidad que no se refiera a los comités, sino tan sólo a las acciones de sus miembros individuales, podría no estar omitiendo verdad alguna. Las verdades sobre los comités son, según esta concepción, sustituibles por verdades sobre los individuos. Parfit cree que deberíamos ser reduccionistas sobre las personas. Se puede, según él, dar cuenta de las verdades sobre las personas en términos de verdades sobre los cuerpos y los estados mentales.

**respuesta «de tipo sí o no»:** Hay preguntas que no admiten una respuesta «de tipo sí o no». Por ejemplo, si preguntamos «¿es calvo?» refiriéndonos a un hombre al que aún le queda un poco de cabello, la pregunta podría carecer de respuesta. Un rasgo importante de las preguntas que no admiten una respuesta «de sí o no» es que, por mucha información fáctica que se reúna, no son susceptibles de ser respondidas. La ausencia de una respuesta definida depende de nuestros conceptos, no del mundo. Parfit piensa que algunas preguntas sobre la identidad personal carecen de una respuesta «de tipo sí o no».

**ser necesario:** Dios es tradicionalmente concebido como un ser necesario. Esto es, resultaría imposible que no existiera. Para decirlo en la jerga de los mundos posibles, existiría en todo mundo posible. Algunos filósofos, como por ejemplo Bertrand Russell, han cuestionado la coherencia de la frase «ser necesario», pero las razones que han aducido son dubitables.

**si, y sólo si:** Una oración de la forma «P si, y sólo si, Q» es equivalente a la conjunción «si P, entonces Q y si Q, entonces P», y es por lo tanto verdadera únicamente si P y Q tienen el mismo valor de verdad.

**sustancia:** Esta noción, tal como es usada aquí, procede de Aristóteles. Las entidades concretas y, en especial, las entidades biológicas (Platón, Sócrates, este caballo, este árbol, etc.) constituyen para Aristóteles las «sustancias primarias». Si el mundo hubiera carecido completamente de sustancias primarias, no contendría nada. Las «sustancias secundarias» tales como la *equinidad* y la rojez serían, por su parte, sustancias dependientes. Si no hubiera caballos o cosas rojas, la *equinidad* y la rojez no existirían. Esto ayuda a explicar la discrepancia de Aristóteles con respecto a Platón sobre la naturaleza de los universales.

# Índice analítico